Cross Layer Design in MIMO Multi-cell Systems

Subhash Lakshminarayana

THÈSE DE DOCTORAT

SPÉCIALITÉ: Télécommunications

École doctorale "Sciences et Technologies de l'Information, des Télécommunications et des Systèmes"

Présentée par :

SUBHASH LAKSHMINARAYANA

Sujet:

Conception de Mecanismes Inter-couches dans les Systemes MIMO Multi-cellulaires

(Cross Layer Design in MIMO Multi-cell Systems)

Soutenue le 6 dec 2012 devant les membres du jury:

M. Pierre Duhamel,	CNRS/Supélec	Examinateur, Président du Jury
M. Eduard Jorswieck,	Dresden University of Technology	Rapporteur
M. Dirk Slock,	Eurecom	Rapporteur
M. Slawomir Stanczak,	Technische Universität Berlin	Examinateur
M. Thomas Bonald,	Telecom ParisTech	Examinateur
M. Mérouane Debbah,	Supélec	Examinateur, Encadrant de Thèse
M. Mohamad Assaad,	Supélec	Examinateur, Encadrant de Thèse

Contents

Abstract

Future wireless communication systems are expected to see an explosion in the wireless traffic which is mainly fueled by mobile video traffic. Due to the time varying and bursty nature of video traffic, wireless systems will see a wider range of fluctuations in their traffic patterns. Therefore, traditional physical layer based algorithms which perform resource allocation under the assumption that the transmitters are always saturated with information bits, might no longer be efficient. It is, thus, important to design dynamic resource allocation algorithms which can incorporate higher layer processes and account for the stochastic nature of the wireless traffic.

The central idea of this thesis is to develop cross-layer design algorithms between the physical and the network layer in a multiple input multiple output (MIMO) multi-cell setup. Specifically, we consider base stations (BSs) equipped with multiple antennas serving multiple single antenna user terminals (UTs) in their respective cells. In contrast to the previous works, we consider the randomness in the arrival of information bits and hence account for the queuing at the BSs. With this setup, we develop various cross-layer based resource allocation algorithms. We incorporate two important design considerations namely decentralized design and energy efficiency. In particular, we focus on developing decentralized beamforming and traffic flow controller design, energy efficient design under time average QoS constraints and decentralized scheduling strategy in a multi-cell scenario. To this end, we use tools from Lyapunov optimization, random matrix theory and stochastic control theory.

Abstract (French)

Les prévisions relatives trafic de données au sein des systèmes de communications sans-fil suggèrent une croissance exponentielle, principalement alimentée par l'essor de transferts vidéo mobiles. Etant donné la nature soudaine et fluctuante des demandes de transfert vidéo, il faut dès à présent réfléchir à de nouveaux algorithmes d'allocation de ressources performants. En effet, les algorithmes en couche physique traditionnels, qui réalisent de l'allocation de ressources sous l'hypothèse classique que les transmetteurs sont toujours saturés avec des bits d'information, risquent à l'avenir de s'avérer inefficients. Pour cette raison, les algorithmes de demain se doivent d'être dynamiques, dans le sens où ils seront capables de prendre en compte la nature stochastique des fluctuations du trafic de données et qu'ils intégreront des informations issus de processus de couches supérieures.

L'idée centrale de cette thèse est de développer des algorithmes, travaillant avec des informations issues de la couche PHY et de la couche NET, dans un scénario Multi-cells et MIMO (Multiple Inputs, Multiple Outputs).Plus particulièrement, nous considérons un réseau de stations de base (BS) équipés avec plusieurs antennes, chargés de servir plusieurs terminaux mobiles équipés d'une seule antenne (UT) dans leurs cellules respectives. Ce qui nous différencie des travaux précédents, c'est que nous tenons compte de l'aléa avec lequel des demandes de transferts peuvent arriver et que, pour cette raison, nous modélisons la formation de queue de données au niveau des stations de base. Dans cette disposition, nous développons plusieurs algorithmes multicouches, réalisant de l'allocation de ressources décentralisée, et ce, dans une optique d'efficacité énergétique. En particulier, il s'agit ici de réaliser des algorithmes réalisant du beamforming de façon décentralisée et capables de contrôler des fluctuations de trafic, des algorithmes optimisant l'efficacité énergétique sous une contrainte de qualité de service moyenne, des algorithmes de planification décentralisés dans des scénarios multi-cellulaires. Dans cette perspective, nous choisissons de recourir non seulement à des outils d'optimisation de la théorie

de Lyapunov, mais également à la théorie des matrices aléatoires et à la théorie du contrôle stochastique.

Résumé

Introduction

Dans ce manuscrit, nous présentons une description générale d'un problème multi-couches dans un contexte MIMO multi-cellules. En particulier, nous identifions les motivations principales et les défis technologiques qui accompagnent ces problèmes classiques. Enfin, nous abordons les contributions scientifiques principales, apportées dans le cadre de cette thèse.

Motivations et défis technologiques

Durant la dernière décennie, la demande pour des transmissions de données sans fil a fortement augmenté et les prévisions suggèrent une augmentation de type exponentielle dans le futur [1]. Dans ce contexte, plusieurs techniques ont été proposées pour répondre à ce besoin croissant, dont deux techniques importantes, abordées lors de cette thèse : la densification du réseau sans fil et des techniques multi-antennes. Nous allons, dans cette partie, aborder ces deux techniques plus en détail.

Une étude des différentes techniques, développées au cours des 50 dernières années pour permettre d'augmenter les débits de transmissions de données dans les réseaux sans fil, montrent que les gains les plus élevés en termes d'efficacité spectrale peuvent être obtenu via une densification du réseau [2], i.e. par une réduction de la taille des cellules. Cela amène, intrinsèquement, au concept de small cells networks (SCNs, réseaux de petites cellules littéralement) [3, 4], qui est principalement basé sur l'idée d'un déploiement dense de stations de base (BS) auto-organisées, peu coûteuses financièrement et énergétiquement, situées plus près des utilisateurs qu'elles doivent servir. La réduction de distance entre les transmetteurs et les récepteurs permettant alors d'obtenir deux avantages : une meilleure qualité de lien de transmission, ainsi qu'une meilleure réutilisation spatiale des ressources spectrales, le tout à une consommation énergétique plus

faible. Cependant, le concept de SCNs apporte son lot de défis technologiques non résolus, tels que le traitement d'interférences, l'organisation du trafic, la disposition optimale des cellules, la gestion du handover dans ce contexte, des questions de sécurité, l'infrastructure backhaul, etc... Pour une étude détaillée de l'état de l'art sur les techniques liées aux SCNs, nous invitons le lecteur à se référer à [3, 4].

D'autre part, les gains potentiellement offerts par de telles techniques sont depuis longtemps admis dans la littérature [5] et sont parfois pour la plupart déjà implémentés à l'heure actuelle [6]. Le concept de MIMO a par exemple déjà été introduit dans plusieurs travaux tels que [7, 8]. Les techniques MIMO offrent des gains en terme de puissance, permettent une qualité de lien plus élevée et permettent également une augmentation significative du débit, en multiplexant plusieurs flux de data dans un seul et même resource block.

Pour ces raisons et puisque ces techniques joueront probablement un rôle majeur dans les réseaux futurs, nous considérerons, dans le cadre de cette thèse, un modèle multi-cellules MIMO.

La majeure partie des travaux récents abordant le concept de MIMO sont basés sur des considérations de couche physique et ont tendance à ignorer des aspects liés aux couches supérieures (par exemple, les dynamiques du trafic, le control des flux, etc...). Les problèmes précédents sont posés comme des problèmes d'optimisation statiques, qui assument que les transmetteurs sont toujours saturés avec des bits d'information à envoyer au récepteur et ont donc tendance à oublier les aspects dynamiques du trafic. Cependant, dans les réseaux cellulaires du futur, les variations ressenties au niveau du trafic peuvent s'avérer très significatives, comme expliqué dans [9, 10]. Ce résultat est principalement dû à la demande croissante et récente de transferts data affiliés à la vidéo, qui varie au cours du temps en bursts difficilement prévisibles. Dans ce scénario, une pré-allocation de ressources simplement basée sur des considérations de couche physique peut aboutir à une sous-utilisation des ressources disponibles. Un exemple parlant s'avère être un scénario dans lequel chaque BS pré-alloue certaines ressources à ses utilisateurs, mais ne possède pas d'information sur leurs réels besoins de transmission (en particulier si leur buffer est vide ou plein à cet instant). Cette sous-utilisation combinée avec le nombre croissant de souscripteurs mobiles implique directement que la capacité offerte par utilisateur soit compromise. Conséquence directe de cela étant que les réseaux pourraient à l'avenir ne plus être capables de satisfaire la demande croissante d'un nombre croissant d'utilisateurs. Pour cette raison, les réseaux cellulaires du futur vont devoir prendre en compte les aspects dynamiques et stochastiques

du trafic dans leurs algorithmes d'allocations. Des gains significatifs peuvent être obtenus en partageant l'information au niveau de plusieurs couches et en réalisant une optimisation multi-couches.

Pour ces raisons, nous considérons dans cette thèse un scénario multi-cells MIMO, qui implémente des aspects dynamiques au niveau du trafic. En particulier, on considère que les schémas représentant les arrivées dans les buffers des BS sont modélisés aléatoirement et on étudie alors la stabilité des queues de paquets, le contrôle des flux de trafic et l'optimisation de l'allocation de ressources dans ce contexte MIMO multicellulaire.

Par le passé, de nombreux travaux ont cherché à traiter du problème de la stabilité des queues de paquets dans les réseaux sans fil. L'idée principale est basée sur les travaux remarquables de Tassiulas et Emphremendis [11], qui ont introduit le résultat de *maximum weighted based scheduling* , garantissant une optimalité en termes de débits. Par la suite, ces travaux ont été généralisés dans différents papiers [12, 13, 14, 15]. Cependant, la majeure partie des travaux mentionnés précédemment simplifient la façon de modéliser la couche physique, rendant souvent l'implémentation en pratique peut réaliste, voire impossible. En particulier, ils modélisent l'interférence en admettant que deux liens ne peuvent transmettre à un même moment si l'un des liens est à portée du second. Il s'agit alors de définir un ensemble de liens susceptibles de transmettre au même moment, sans générer de conflits. Cependant un tel modèle suppose que l'on considère un schéma de transmission où les canaux sont orthogonalement répartis. Les récentes avancées dans les domaines de power control, beamforming et traitement d'interférence permettent aujourd'hui à plusieurs utilisateurs de transmettre en utilisant le même resource block. Dans ce contexte, il semble important de considérer des aspects stochastiques au niveau de la façon dont sont modélisés les trafics. C'est l'enjeu principal de cette thèse, à savoir être capable de prendre en compte des aspects aléatoires de modélisation du trafic dans des modèles technologiques récents de réseaux MIMO multicellulaires.

Les designs d'algorithmes proposés dans cette thèse traitent également deux problèmes d'égale importance, souvent référés comme design décentralisé et efficacité énergétique.

- *Design décentralisé multi-couches:* Etant donné le nombre massif de BS déployées dans des modèles SCNs, une entité centralisée qui gère la coordination des différentes BS est pour l'instant improbable. Pour rendre ce design implémentable en pratique aujourd'hui, les SCNs doivent fonctionner de façon décentralisée. Cela suppose donc que chaque cellule doit mettre en pratique son processus d'optimisation en utilisant l'information

disponible localement, ainsi qu'un nombre limité d'informations obtenues via un échange avec d'autres cellules distantes. Au même moment, nous devons nous assurer qu'un processus décentralisé ne détériore pas trop la performance globale de notre système. Pour cette raison une question que nous posons dans cette thèse porte sur l'optimisation décentralisée des SCNs. En particulier, nous formulerons nos algorithmes d'optimisation de telle sorte à obtenir une bonne performance en limitant les échanges d'information nécessaires entre des cellules distantes.

- *Efficacité énergétique de notre algorithme:* L'explosion du trafic est susceptible d'être accompagné d'une augmentation drastique de la consommation énergétique au niveau de notre réseau, qui peut être interprété en termes d'indice carbone [16]. Créer un design énergétiquement efficace est donc important dans le sens où il réduira alors les coûts de fonctionnement associés aux SCNs [17]. Il s'agit donc d'être capable de créer des designs dans lesquels on cherche à minimiser la consommation énergétique du réseau en maintenant une qualité de service (QoS) adéquate pour les utilisateurs, le tout, dans un contexte mulitcellulaire MIMO.

Nous allons à présent détailler l'organisation de cette thèse et les principales contributions apportées.

Organisation et contributions

Le modèle suivant, détaillé en figure 1, est celui que nous avons considéré dans l'intégralité des scénarios abordés dans cette thèse.

Ce modèle représente un scénario multicellulaire, dans lequel chaque BS doit servir ses utilisateurs dans ses cellules respectives. Chaque BS est par ailleurs équipée d'antennes multiples et chaque UT ne possède qu'une seule antenne. Les BS possèdes des queues dédiées utilisées comme buffers pour stocker les paquets des utilisateurs qu'elles sont supposées servir. Le trafic est modélisé par des arrivées de paquets au niveau du réseau, directement intégrées dans les queues de chaque BS, afin de pouvoir servir les utilisateurs plus tard. Les BS ont un objectif consistant à servir les utilisateurs en partageant les ressources allouées à leurs utilisateurs de telle sorte à stabiliser leurs queues. Nous considérons donc le problème de stabilisation de queues, basées sur les hypothèses précédentes et basés sur différentes connaissances possibles de l'état du canal (Channel State Information, CSI). De plus, étant donné que les échanges d'information entre BSs est coûteux, nous formulons des algorithmes d'allocation qui considèrent

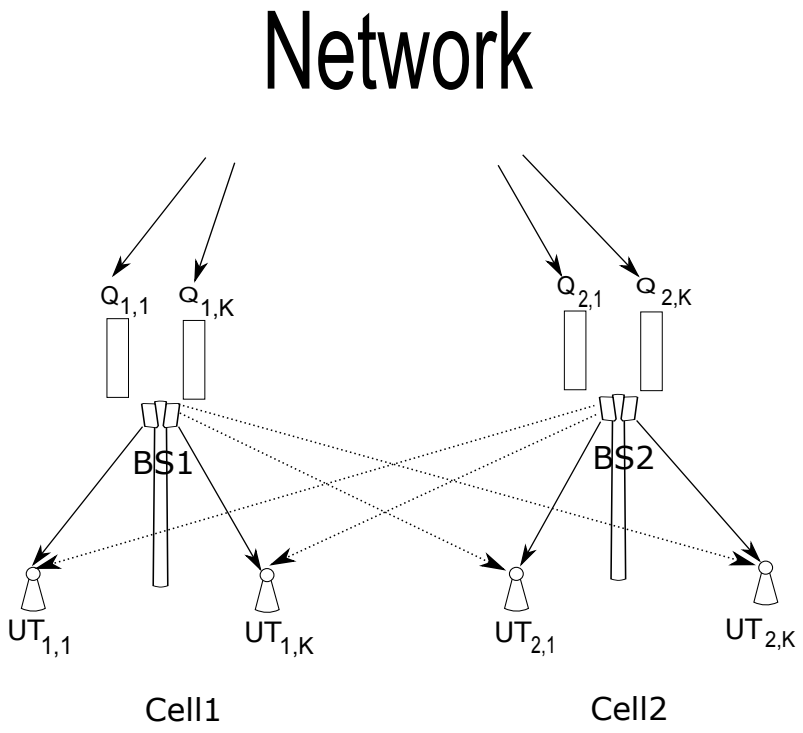

Figure 1: Schéma de modèle cohérent dans un contexte multicellulaire MIMO. Chaque BS est alors équipée de queues utilisées comme buffers pour stocker les paquets des utilisateurs qu'elles doivent servir.

des échanges d'information possibles, mais en nombre limité.

Etant donné la formulation de modèle précédente, nous pouvons alors poser différents problèmes d'allocation de ressources et étudier diverses solutions décentralisées.

La thèse est alors divisée en trois parties. Premièrement, dans le chapitre 2, nous présentons un état de l'art des résultats déjà disponibles pour des problèmes similaires, qui seront utilisés dans cette thèse. Dans le chapitre 3, nous traitons le cas d'un problème joint de contrôle de trafic et de traitement d'interférence dans des réseaux multicellulaires MIMO. Nous considérons un contrôleur de flux connecté à un grand nombre de BS, dont la mission est de réguler le trafic et les arrivées dans les queues de chaque BS. Le trafic doit ensuite être transmis via une transmission sans fil d'une BS vers l'utilisateur approprié. Les BSs réalisent pour cela du beamforming coordonné multicellulaire MIMO, afin de transmettre efficacement leurs données. Le nombre de bits envoyés et liquidés dans la queue de chaque BS dépend alors du SINR entre la BS et l'utilisateur considéré. Ce qui amène à prendre en compte deux aspects alors essentiels :

- Le contrôleur de flux doit alors ajuster ses décisions de régulation en prenant en compte la connaissance partielle ou complète à propos du CSI relative aux liens sans fils entre les BSs et les utilisateurs.
- Les BSs doivent réaliser un beamforming coordonné multicellulaire, en limitant les échanges d'informations entre BSs.

La première hypothèse repose principalement sur le fait que le contrôleur de flux est connecté à un grand nombre de small cells BSs and ne peut tenir compte de toutes les conditions et évolutions des canaux au niveau de toutes les BSs. Le deuxième permet au design SCNs de pouvoir être implémenté sans nécessiter pour autant une infrastructure de backhaul trop importante.

Le chapitre 3 est divisé en deux parties. Dans la première partie, on développe un modèle réduit de beamforming multicellulaire. Dans lequel les BSs n'ont besoin d'échanger que les statistiques associées aux dynamiques des canaux entre eux, plutôt qu'un CSI instantané. Théoriquement, la performance d'un tel algorithme correspond exactement à celle acquise par un algorithme centralisé (avec un échange en temps réel des CSI instantanés), lorsque le nombre de BSs et d'utilisateurs augmente à l'infini. Pour cette raison, nous décidons donc de recourir à des outils de théorie des matrices aléatoires (Random Matrix Theory, RMT). Dans la seconde partie, nous nous interessons à un problème de contrôle de flux couplé avec un design de beamforming décentralisé comparable à celui décrit dans la partie précédente. Plus particulièrement, on développe dans ce

cadre, un contrôleur de flux H^∞ capable de réguler le trafic au niveau des BSs et conscient des dynamiques liées aux CSI entre les BSs et les utilisateurs. Les résultats de cette section ont été publiés dans les articles suivants :

[J1] S. Lakshminaryana, J. Hoydis, M. Debbah and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," submitted to *IEEE Transactions on Signal Processing,* 2012

[J2] S. Lakshminarayana, M. Assaad and M. Debbah, "H^∞ Control Based Scheduler for the Deployment of Small Cell Networks," to appear in *Elsevier Performance Evaluation Journal*, Special Issue of selected papers from WiOpt 2011 (accepted for publication)

[C1] S. Lakshminarayana, M. Assaad and M. Debbah, "H^∞ Control Based Scheduler for the Deployment of Small Cell Networks," *9th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'11)* , Princeton, USA, 2011

[C2] S. Lakshminaryana, J. Hoydis, M. Debbah and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC'10)*, Istanbul, Turkey, 2010

Dans le chapitre 4, nous nous intéressons plus spécifiquement à la conception d'algorithmes énergétiquement efficients, via la question : Dans un réseau multicellulaire MIMO, étant donné une contrainte moyenne de qualité de service (QoS) pour les utilisateurs, quelle est l'énergie minimale permettant de satisfaire cette contrainte ? La contrainte de QoS ici considérée est la différence entre la moyenne de la puissance de signal moyenne par unité de temps et la moyenne de la puissance de l'interférence par unité de temps, que l'on contraint d'être supérieur à une valeur prédéfinie. L'intérêt principal de considérer des critères de QoS moyennés par unité de temps repose sur le fait qu'il apporte de la flexibilité pour allouer dynamiquement nos ressources sur différents canaux à fading dont on connait les statistiques, comparé à des designs où l'on posséderait une connaissance instantanée des contraintes de QoS. Il peut également aboutir à une réduction en moyenne de la consommation énergétique de notre réseau. D'un point de vue pratique, l'intérêt principal de considérer des contraintes de QoS moyennes repose sur le fait que des applications classique de services de données (partage de fichier, téléchargement vidéo,...) permet certaines flexibilités et accepte facilement certains délais.

On définit, tout d'abord, l'ensemble faisable, satisfaisant la contrainte de QoS, par l'ensemble de toutes les contraintes de QoS moyennes pouvant être atteintes par une stratégie d'allocation. Pour cela, on modélise les contraintes comme des queues et on considère alors le problème consistant à stabiliser ces queues, tout en minimisant la consommation énergétique du réseau. L'optimisation utilise une optimisation basée sur la technique de Lyapunov [13], qui permet de formuler un problème de contrôle dynamique, visant à minimiser la consommation énergétique. L'approche considérée met en œuvre une série d'optimisations statiques qui doivent être résolus pour chaque période de temps. Ce problème peut être reformulé comme un problème semi-definite programming (SDG). A chaque instant, la solution au problème d'optimisation offre le vecteur de beamforming optimal à chaque instant. Notre algorithme permet alors de considérer un design décentralisé dans lequel les BSs peuvent reconstruire et approcher le vecteur de beamforming optimal en utilisant une information locale réduite à propos du CSI local. L'échange d'information entre BS se résumerait alors à un échange d'information sur leurs longueurs de queues respectives. Dans ce contexte, nous apportons une comparaison de performance entre notre algorithme décentralisé et le problème centralisé avec connaissance totale : on montre alors que l'écart de consommation énergétique entre la stratégie décentralisée et la stratégie optimale sont acceptables. Par la suite, nous caractérisons la performance de nos algorithmes dans des cas où l'information échangée s'avère être délayée : il s'avère que la contrainte de QoS peut également être atteinte, mais que le délai affecte cependant la performance en terme d'efficacité énergétique, la réduisant jusqu'à une valeur constante.

Les résultats de cette question peuvent être trouvés dans les articles suivants :

[J3] S. Lakshminaryana, M. Assaad and M. Debbah, "Energy Efficient Cross-Layer Design in MIMO Multi-cell Systems," **to be submitted**

[C3] S. Lakshminaryana, M. Assaad and M. Debbah, "Energy Efficient Cross-Layer Design in MIMO Multi-cell Systems," **to be submitted** International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13), 2013

Dans le chapitre 5, nous considérons un problème d'allocation décentralisé dans un scénario consistant en deux BSs devant servir leur utilisateur respectif (un utilisateur pour une BS). L'objective dans ce contexte est de stabiliser les queues pour un flux de trafic en arrivée au niveau des queues connu. On admet que ce flux de trafic permet de garantir la stabilité des queues rendant le

problème réalisable. Nous proposons, ainsi, un algorithme appelé Fast CSMA (Carrier Sense Multiple Access) basé sur un algorithme décentralisé, modélisant l'interférence via un critère de SINR. Dans notre algorithme, les BSs, doivent seulement échanger un bit de données entre elles. D'une part, on peut remarquer que l'application directe de l'algorithme FCSMA dans un modèle à interférence basée sur le SINR ne permet pas d'atteindre une bonne performance. Cependant, en combinant cet algorithme avec une règle de séparation dynamique du trafic, la performance obtenue redevient correcte. L'aspect innovant de cet algorithme repose alors sur la nature décentralisée de l'opération et la possibilité de l'appliquer à des scénarios de canaux à évanouissement.

Les résultats de cette section peuvent être trouvés dans l'article suivant :

[C2] S. Lakshminarayana, B. Li, M. Assaad, A. Eryilmaz and M. Debbah, "A Fast-CSMA Based Distributed Scheduling Algorithm under SINR Model" *IEEE International Symposium on Information Theory (ISIT'12)*, Cambrigde, MA, USA, 2012

La thèse se conclue avec un chapitre 6, qui résume quelques-uns des résultats majeurs de cette thèse et propose un aperçu des travaux futurs.

Autres publications réalisées dans le cadre de cette thèse :

[C3] S. Lakshminarayana, M. Debbah and M. Assaad, "Asymptotic Analysis of Downlink Multi-cell Systems with Partial CSIT", *IEEE International Symposium on Information Theory (ISIT'11)* , St-Petersburg, Russia, 2011

Conclusion et travaux futurs

Les prévisions envisagent une explosion du trafic de données au sein des réseaux sans fils du futur, ce qui aura pour conséquence d'accroitre de façon significative la pression générée dans la gestion de la capacité au sein des réseaux déjà existants. Pour cette raison, de nombreux chercheurs du domaine académique, autant que du domaine industriel, ont cherché ces dernières années à élaborer des techniques, qui permettront au réseau actuel d'encaisser au mieux cette explosion de la demande.

Cette thèse se concentre principalement sur un contexte multicellulaire MIMO. La plupart des algorithmes déjà existants, cherchent à optimiser des critères purement affiliés à des considérations de couche physique. Ils réalisent une allocation de ressources, de façon statique et ne prennent pas en compte la nature stochastique du trafic. Dans la même période, des travaux se sont attachés à

étudier des problèmes de stabilité de queues au niveau des réseaux. Cependant, dans la majorité de ces travaux, la modélisation au niveau de la couche physique est simplifiée, rendant le modèle peu réaliste et difficilement implémentable en pratique. Pour cette raison, il est nécessaire de concevoir de nouveaux algorithmes qui peuvent prendre en compte à la fois la nature stochastique du trafic, mais également des aspects réalistes en terme de couche physique, tels que le fading, le path loss, l'interférence, une connaissance imparfaite du CSI, du contrôle de puissance, des technologies MIMO,etc... Cette thèse propose quelques algorithmes prenant en compte ces deux considérations, en gardant à l'esprit que les algorithmes devront fonctionner de façon décentralisée et devront être énergétiquement efficaces.

Nous résumons, à présent, brièvement les résultats principaux de cette thèse et mettons en évidence les possibles limites et/ou extension de notre modèle et des algorithmes qui y sont associés.

Dans le chapitre 3, nous avons formulé un algorithme de beamforming avec un overhead de taille réduite, en employant des outils de la théorie des matrices aléatoires. En utilisant une analyse asymptotique, nous avons prouvé l'optimalité de la stratégie de beamforming à overhead réduit. Dans un scénario MIMO, nous avons également mis en évidence les sets de stratégies faisables en termes de SINR, dans un réseau cellulaire MIMO, en utilisant des outils d'analyse de systèmes larges. Nos résultats montrent qu'un échange d'information à une échelle de temps proche de celle des dynamiques du canal permet d'atteindre de bonne performance en pratique. Une amélioration envisagée consiste à caractériser les fluctuations en termes de SINR perçu en downlink, dans un système à dimensions finies. Pour tenir compte des fluctuations du SINR autour de la valeur visée, il est possible de résoudre le problème en utilisant un algorithme ROBF et en considérant une valeur plus élevée pour le SINR que la valeur précédemment voulue.

Plus tard dans ce chapitre, nous développons un contrôleur de flux pour systèmes MIMO multicellulaires en utilisant un contrôle H^{∞}, qui peut fonctionner sans connaissance des CSI associés aux différents liens sans fils du système. Nous voudrions souligner que dans notre solution, nous avons découplé les problèmes de beamforming et de contrôle de flux en deux parties distinctes. Cependant, une méthode meilleure consisterait à réaliser l'optimisation conjointe du contrôleur de flux et du problème de beamforming (comme par exemple dans [13], qui apporte un exemple d'algorithme optimisant à la fois le problème de beamforming et le contrôleur de flux, dans un réseau SISO et un modèle d'interférence conflictuel basé sur un graphe). Cette distinction peut

être perçue dans notre travail comme un sacrifice volontaire de performance, permettant de rechercher alors une solution au problème équivalent décentralisé.

Dans le chapitre 4, nous avons considéré le problème visant à minimiser la consommation énergétique d'un réseau soumis à une contrainte de QoS à long terme dans un contexte de réseau MIMO multicellulaire. Nous avons montré que des contraintes de QoS moyennes permettent d'atteindre de meilleurs résultats en termes d'efficacité énergétique, comparé au problème équivalent dans lequel les contraintes de QoS seraient données instantanément à chaque instant. Une future direction possible pourrait consister à trouver le vecteur de beamforming optimal associé à un problème de transmission devant satisfaire une contrainte de QoS moyennée en temps (et de la forme log(1+SINR)) et renvoie ainsi à un problème réaliste de stabilisation de queue. L'application de technique d'optimisation Lyapunov à ce problème aboutit à un problème de maximisation de débit moyen, dans un scénario MIMO multicellulaire, connu pour être un problème non-convexe. Une extension également possible consisterait à réaliser l'optimisation conjointe appliquée au feedback CSI (optimiser la formation de canal durant chaque time slot et optimiser le nombre de bits alloués au feedback dans le cas d'une quantification pour le feedback) et du problème de beamforming.

Finalement, dans le chapitre 5, nous avons développé un algorithme d'allocation distribué basé sur la technologie FCSMA, dans un scénario de canaux interférés et une interférence basée sur le SINR. Pour des raisons de malléabilité, nous avons considéré un modèle symétrique. Des généralisations possibles de cet algorithme FCSMA dans un cas général de réseau ont été mentionnées à la fin du chapitre 5.

Nous apportons finalement des conclusions générales dérivées de cette thèse. Durant cette thèse, nous avons tenté d'incorporer des considérations réalistes telles que des flux de trafic et des gestions de queues dans des systèmes multicellulaires MIMO. D'un point de vue couche physique, nos travaux apportent un aperçu basique de la façon dont il est possible de gérer l'aspect stochastique du trafic et de maintenir une stabilité acceptable au niveau des queues de systèmes MIMO. Nous montrons que les améliorations en termes de performance sont obtenues en réalisant une optimisation de notre système de façon multi-couche, plutôt que de réaliser une optimisation classiquement purement basée sur la couche physique (nos résultats montrent une amélioration en termes d'efficacité énergétique). Nos travaux contribuent également à prendre en compte des aspects de décentralisation au niveau de nos algorithmes d'optimisation, qui permettent d'atteindre des performances sensiblement équivalentes à celles obtenus

avec des optimisations centralisées. Nos travaux soulèvent finalement des questions ouvertes sur la façon dont l'optimisation de problèmes purement associés à la couche physique (tels que le fading, le path loss l'interférence, une connaissance imparfaite du CSI, technologies MIMO...) peuvent être géré au niveau du réseau.

Nous concluons finalement cette thèse en donnant quelques idées possible de poursuites de travaux.

Travaux futurs

Feedback CSI et stabilité de queues dans un réseau MIMO multicellulaire

Une des limitations fondamentales dans la performance des systèmes sans fils consiste en l'cquisition du CSI. En particulier, les schémas de feedback CSI ont longtemps été étudiés dans le contexte multi-utilisateur MIMO [18, 19] (feedback numérique et analogique) avec l'objectif de maximiser certaines métriques telles que le débit moyen ergodique. Dans cette thèse, nous avons essayé de mettre en relation la notion de stabilité de queues avec des techniques de transmission utilisant des antennes multiples. Un extension évidente de nos travaux consisterait à mettre en relation le problème de stabilité de queues avec celui de l'acquisition d'un feedback CSI. Les questions suivantes se posent alors. Comment évolue la région de stabilité des queues avec le taux moyen de feedback CSI ? Etant donné un débit moyen de feedback CSI, quand et avec quelle précision les UTs peuvent-ils transmettre leur feedback CSI aux BSs ?

Un autre problème intéressant consisterait à concevoir un système gérant conjointement les feedbacks CSI et les techniques de transmissions, capable d'une certaine performance avec un nombre de feedbacks CSI réduits. Dans ce but, une direction intéressante serait d'exploiter la corrélation temporelle inhérente aux processus stochastiques associés aux canaux. La plupart des travaux utilisent des processus i.i.d. pour modéliser l'évolution des canaux au cours du temps. Ce modèle a été un choix populaire en raison de sa simplicité, permettant de réduire drastiquement la complexité de l'analyse des systèmes dans lequel il est considéré. Mais, ce modèle est insuffisant, car il ne permet pas de prendre en compte la corrélation temporelle inhérente aux canaux. On peut, cependant, modéliser ce phénomène, en recourant à des canaux de Markov [20]. La mémoire inhérente aux canaux peut aider à réduire le besoin de feedbacks CSI, comparé à des modèles i.i.d. Le problème d'optimisation conjointe

peut alors être modélisé et résolu via des des processus de décision Markoviens (littéralement, Markov Decision Processes, MDPs), comme explicité dans les travaux [21, 22].Un autre scénario intéressant consiste à optimiser le feedback CSI en considérant un modèle de canaux qui admet un corrélation temporelle mais est non-stationnaire, qui peut être acquis via des techniques telles que des algorithmes d'apprentissage.

Algorithmes efficaces en termes de délais

Les nouvelles applications telles que le vidéo streaming, les appels vidéos et les jeux en ligne demandent de bonnes garanties en termes de performance de délais. La conception d'algorithmes capables de contenir le délai dans des proportions acceptables, pour des réseaux sans fils a sollicité énormément d'attention ces dernières années [23, 24]. Il est bien connu que l'utilisation d'antennes multiples apporte des degrés de liberté supplémentaires. Une question intéressante serait alors de savoir comment exploiter ces degrés supplémentaires pour offrir une meilleure performance en termes de délais. Il est également possible de poser le problème de l'exploitation d'antennes multiples pour gérer un trafic sous une contrainte de délai (ce qui est primordial dans certaines applications multimédia).

Stratégies d'allocation opportunistes pour des réseaux cognitifs hétérogènes

Les idées développées dans cette thèse peuvent être étendues aux scénarios de réseaux cognitifs hétérogènes [25, 26, 27]. La notion de réseaux cognitifs hétérogènes se rapporte à un réseau sans fils classique auquel on superpose une densification d'autres cellules (par exemple des femtocellules), capable de prendre en compte les variations de son environnement de façon intelligente et de réaliser, en conséquence, une allocation de ressources dynamique. Dans de tels réseaux, on espère être capable de gérer l'allocation de ressource entre les femtocellules et les macrocellules , grâce à une adaptation dynamique du réseau, par rapport aux changement ressentis de son environnement. Dans un tel scénario, on peut alors se poser les questions suivantes. Etant donné une contrainte de QoS pour les utilisateurs associés à des macrocellules, quel est le niveau maximal d'utilité pouvant être atteint et garantissant la coexistence des transmetteurs-récepteurs cognitifs ? De façon équivalente, combien de liens cognitifs transmetteurs-récepteurs peuvent être actifs à un même instant, sans perturber les transmissions des utilisateurs associés aux macrocellules ? Ces

problèmes peuvent être formulés comme des problèmes d'optimisation stochastiques et il est possible de recourir aux outils d'optimisation de Lyapunov, afin d'en déduire des algorithmes efficaces dans ce contexte. Ce scénario s'avère particulièrement intéressant, dans le sens où les aspects symmétriques des flux de données des utilisateurs macrocellulaires, l'aspect stochastique du trafic, et les états des différents canaux, vont offrir un multiplexage statistique, qui constituera une opportunité permettant aux utilisateurs cognitifs de coexister. Les paires de transmetteurs-récepteurs cognitifs peuvent ainsi utiliser de façon opportuniste les ressources disponibles, en garantissant à chacun une bonne qualité de service.

Notations

$\mathbb{C}^{N\times M}$	the set of complex-valued $N \times M$ matrices
$\mathbb{R}^{N\times M}$	the set of real-valued $N \times M$ matrices
$\mathbb{C}^N, \mathbb{R}^N$	short form for $\mathbb{C}^{N\times 1}$ and $\mathbb{R}^{N\times 1}$ (column vectors)
$\mathbf{X}$	matrix
$\mathbf{X}(i,j)$	(i,j) entry of $\mathbf{X}$
$\mathrm{tr}(\mathbf{X})$	trace of the matrix $\mathbf{X}$
$\mathrm{rank}(\mathbf{X})$	rank of the matrix $\mathbf{X}$
$\mathbf{X}^T$	transpose of the matrix $\mathbf{X}$
$\mathbf{X}^H$	complex conjugate transpose of the matrix $\mathbf{X}$
$\mathbf{I}_M$	identity matrix of size M
$\mathrm{diag}(x_1, ..., x_M)$	diagonal matrix with the elements x_i along its main diagonal
$\mathcal{CN}(\mathbf{m}, \mathbf{R})$	complex Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$
$\xrightarrow{\text{a.s.}}$	almost sure convergence
$\mathbb{E}[X]$	expectation of a random variable X
$a_N \asymp b_N$	$a_N - b_N \xrightarrow{\text{a.s.}} 0$
$\mathbf{x}$	column vector
x_i	i^{th} element of the column vector
$\mathbb{C}, \mathbb{R}$	the spaces of complex and real numbers
$(x)^+$	$\max(x, 0)$
$\rho(\mathbf{X})$	spectral radius of matrix $\mathbf{X}$
$\mathbf{X} \leq \mathbf{Y}$	element wise inequality
$\log(x)$	natural logarithm
$N_t, K \to \infty$	$N_t, K \in \mathbb{N}^+$, the following condition on N_t and K, $0 < \liminf_{K\to\infty} \frac{N_t}{K} \leq \limsup_{K\to\infty} \frac{N_t}{K} < \infty$.
$\mathcal{S} \setminus x$	the remaining set when member x is removed
$\mathbf{0}$	vector of zeros

Acronyms

BS	base station
CS	central station
CSI	channel state information
CSMA	carrier sense multiple access
CTMC	continuous time Markov chain
FCSMA	fast CSMA
e.s.d.	empirical spectral distribution
ICT	information and communication technology
i.i.d.	independent and identically distributed
LHS	left hand side
LQG	linear quadratic Gaussian
MAC	multiple access channel
MIMO	multiple input multiple output
MISO	multiple input single output
MMSE	minimum mean square error
QoS	quality of service
OFDM	orthogonal frequency division multiplexing
RHS	right hand side
RMT	random matrix theory
Rx	receiver
SCN	small cell networks
SNR	signal to noise ratio
SINR	signal to interference noise ratio
SOCP	second order conic programming
SDP	semi-definite programming
Tx	transmitter
UT	user terminal

Chapter 1

Introduction

In this chapter, we present a general description of the problem of cross-layer design in multiple input multiple output (MIMO) multi-cell systems. In particular, we identify the main motivation and the technological challenges behind this problem. Finally, we describe the main contributions presented in this thesis.

1.1 Motivation and Technological Challenges

During the last decade, there has been an ever increasing demand for wireless data services and it is expected to increase exponentially in the future as well [1]. While many different techniques have been proposed to cope up with the increasing data rate demand, the two important techniques which are relevant to this thesis are the concepts of network densification and multiple antenna techniques. We will now describe these two factors in greater detail.

A study of various techniques that have been developed over the last 50 years to enable higher data rates in wireless communication systems shows that the biggest gains in the efficiency of spectrum utilization has been due to network densification [2], i.e., shrinking of cell sizes. This leads to the concept of small cell networks (SCNs) [3, 4] which is based on the idea of a very dense deployment of self-organizing, low-cost, low-power base stations (BSs), bringing them closer to the user terminals (UTs) they serve. The reduction in distance between the transmitter and the receiver creates the dual benefit of higher link qualities and better spatial reuse. It also implies reduction in the amount of power transmitted. However, the reduction of cell sizes imposes numerous technical challenges in the design of such SCNs. Some of the technical issues faced in the implementation of SCNs are the problems of interference management, traffic

scheduling, optimal cell size planning, call handover issues, security, backhaul infrastructure etc. For a detailed survey on the SCN deployment and design challenges in SCNs, the reader can refer to [3, 4].

On the other hand, the gains of using multiple antennas at the transmitter and receiver have been well understood in literature [5] and has already become a part of the current day cellular systems [6]. The concept of MIMO systems was first introduced in the seminal works of [7, 8]. MIMO techniques can provide power gains, improve the link reliability, and increase the throughput by multiplexing several independent data streams in the same resource block.

The aforementioned factors motivate us to consider in this thesis, a MIMO multi-cell setup as they are expected to play a key role in the design of next generation cellular systems.

Most of the prior works on the design of MIMO systems are based on physical layer considerations and tend to ignore the higher layer process (e.g. traffic dynamics, flow control etc). The problems are posed as static optimization problems which assume that the transmitter is always saturated with information bits to be sent to the receiver and tend to ignore the stochastic nature in the traffic patterns. However, future cellular networks are expected to experience a much wider range of fluctuations in their traffic patterns [9, 10]. This is mainly a result of the increased multimedia related traffic which tends to be time varying and bursty nature. In this scenario, pre-allocation of resources based only on physical layer considerations can lead to underutilization of resources. For example, consider a scenario in which the BS pre-allocates certain resources to a UT during a given time slot but has no information to transmit (because it user's buffer might be empty). The underutilization combined with the increasing number of mobile subscribers implies that the capacity of the present day wireless networks is significantly stressed. As result, cellular networks designed with traditional physical layer optimization based approaches might not be able to support the demands of the large number of UTs. Therefore, future cellular networks require dynamic resource allocation algorithms, which can take into account the stochastic nature of the traffic patterns. Significant gains can be obtained by sharing information across the layers and optimizing the design from a cross-layer perspective.

Motivated by the aforementioned considerations, in this thesis, we consider the design aspects in MIMO multi-cell scenario while incorporating the dynamic nature of cellular traffic. In particular, we consider the randomness in traffic arrival patterns and hence account for queuing of information bits at the BS. We deal with issues such as queuing stability, traffic flow control and cross-layer

based optimization of the system parameters in MIMO multi-cell networks.

There have been many works in the past which deal with the issue of queuing stability in wireless networks. The main idea is based on the seminal work of Tassiulas and Emphremedis [11], which introduced the celebrated result of maximum weight based scheduling with throughput optimality[1] guarantees. Subsequently, this theory has been generalized in many different works [12, 13, 14, 15]. However, most of the aforementioned works simplify the modeling of the physical layer, often making it unrealistic for practical implementation. In particular, they assume a *conflict graph* based interference model at the physical layer, in which two links cannot transmit simultaneously if one link is within the transmission range of the other. The task is then to schedule a set of non-conflicting links for transmission. However, the conflict graph based interference model can only account for orthogonal channel access schemes. Advances in physical layer techniques like multiple antenna based technologies, power control, beamforming and interference alignment allow cellular UTs to share and coexist over the same time/frequency resource blocks. Hence, it is important to consider the randomness of traffic patterns with advanced physical layer models. This is indeed the main idea of this thesis. The work in this thesis attempts to develop a cross-layer design framework in MIMO multi-cell networks.

The design algorithms developed in this thesis are coupled with two issues which are of paramount importance in the design of future cellular networks, namely decentralized design and energy efficiency.

- *Issue of decentralized cross-layer design:* Owing to the sheer number of SCN BSs which are planned to be deployed, a centralized entity coordinating all the SCN transmissions would be impossible. To make the design practically viable, the SCNs must function in a decentralized manner. This implies that each cell should optimize its performance relying mainly on the locally available information and limited amount of additional side information exchanged between the neighboring cells. At the same time, it must be ensured that the decentralized operation of the cells does not significantly deteriorate the overall system performance. Thus, one of the key question we address in this thesis is the decentralized design of SCNs. In particular, we formulate our cross-layer design algorithms with reduced information exchange between the cells which provide a good overall system performance.

- *Issue of energy efficient cross-layer design:* The second concern addressed

[1] A scheduling policy is said to be throughput optimal if it achieves any throughput (subject to network stability) that is achievable by any other scheduling strategy.

in this thesis is the problem of energy-efficient transmissions. The explosion in wireless traffic is likely to cause a substantial increase in information and communication technology (ICT) related carbon emissions [16]. Energy efficient deisgn is also important because it can lead to significant reduction in the cost of operating the network [17]. Therefore, the second important aspect we consider in our cross-layer design algorithms is the issue of minimizing energy expenditure subject to quality of service (QoS) constraints for the users, in MIMO multi-cell networks.

We will now proceed to list the basic outline and the main contributions of the thesis.

1.2 Outline and Contributions

The system set-up which is common to all the scenarios considered in this thesis is provided in Figure 1.1.

The set up consists of a multi-cell scenario. Each cell has a BS serving the UTs in its respective cell. The BSs are equipped with multiple antennas and the UTs have a single antenna each. The BSs have dedicated queues to buffer the packets of the UTs they serve. The traffic arrives from the network into the queues present at the BS, where they can be buffered before being transmitted over the wireless medium. The BSs have to perform the task of resource sharing among the UTs with the objective of stabilizing the queues. We consider the queue-length stabilization problem under various assumptions on channel state information (CSI) knowledge at the BS. Additionally, since information exchange between the BSs is costly, we formulate the resource allocation algorithms with limited information exchange between them.

With the above system set up, we formulate a number of resource allocation problems and provide decentralized solutions. The thesis is divided into three main parts.

First, in Chapter 2, we present some background of the existing results which will be used in this thesis.

In Chapter 3, we deal with the joint problem of traffic flow control and interference management in MIMO multi-cell networks. We consider a flow controller connected to a large number of BSs, which regulates the traffic arrival into the queues present at the BSs. The traffic in turn has to be transmitted over the wireless medium. The BSs perform coordinated multi-cell beamforming in order to efficiently transmit their data over the wireless medium. The number of bits departing from the queues is a function of the achieved SINR. The two

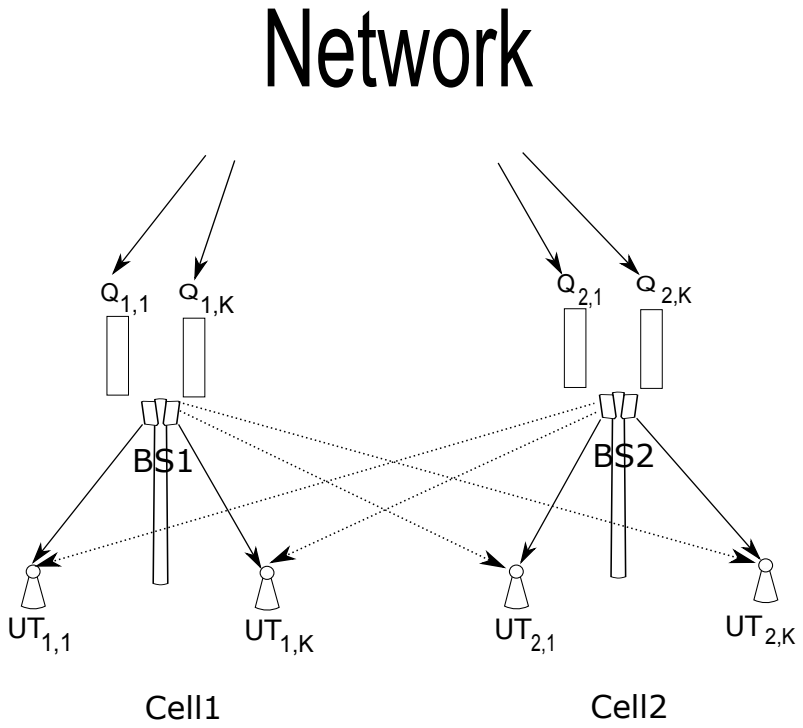

Figure 1.1: System setup consisting of a multi-cell scenario. Every BS is equipped with queues to buffer the packets of the UTs they serve.

important design considerations in Chapter 3 are the following:

- The flow controller must take the flow control decisions in order to ensure the stability of the queues while being oblivious to the CSI of the wireless link between the BSs and the UTs.
- The BSs must perform the joint multi-cell beamforming with limited information exchange between them.

The first design consideration arises from the fact that the flow controller is typically connected to a large number of small cell BSs and hence cannot keep track of the channel conditions of all the BSs. The second design consideration is important in order to enable SCNs to be operational with limited backhaul infrastructure.

Chapter 3 is divided into two parts. In the first part, we develop a reduced overhead multi-cell beamforming design algorithm in which the BSs need to exchange only the channel statistics among themselves rather than the instantaneous CSI. We theoretically prove that the performance of such an algorithm exactly matches that of a centralized algorithm (with instantaneous CSI exchange), when the number of antennas at the BSs and number of UTs grow large. To this end, we use tools from random matrix theory (RMT). In the second part, we address the problem of traffic flow control coupled with the decentralized beamforming design developed in the previous part. Specifically, we develop a H^{∞} control based flow controller which regulates the traffic arrival into the queues present at the BS while being oblivious to the wireless CSI between the BS and the UTs.

In Chapter 4, we address the following question. In a MIMO multi-cell system, given certain time average quality of service (QoS) constraints for the users, what is the minimum energy required to satisfy them. We first characterize the feasible QoS region, which is the set of all time average QoS constraints that can be successfully met by an appropriate policy. We then formulate the problem as a stochastic optimization problem and use tools from Lyapunov optimization to develop a dynamic beamforming solution. The above approach leads to a series of static optimization problems which must be solved during each time slot. We reformulate the static optimization problem as a semi-definite programming (SDP) problem. The solution to the optimization problem provides the optimal beamforming vector design at each instant. Our formulation naturally leads to the decentralized design in which the BSs can formulate the optimal beamforming vectors using only the locally available CSI. The BSs would only have to exchange the limited information among them. We incorporate the case of

delayed information exchange and examine its performance with respect to the optimal solution.

In Chapter 5, we consider a decentralized scheduling problem in scenario consisting of two BSs serving their respective UTs (one UT per BS). The design objective in this case is to stabilize the queues for a given traffic arrival rate (which lies inside the stability region). Specifically, we develop an algorithm called the *Fast CSMA* (carrier sense multiple access) based distributed scheduling algorithm under an SINR based interference model. In our algorithm, the BSs only have to exchange one bit information among them. First, we note the direct application of FCSMA algorithm to an SINR based interference model achieves low performance. We then combine the FCSMA algorithm with a *dynamic traffic splitting rule* and charectarize the performance of our algorithm. The novelty of our algorithm lies in the decentralized nature of the operation and its applicability to fading channel scenario.

The thesis is concluded in Chapter 6 which summarizes some of the main results and provides an outlook to future work.

1.3 Publications

The following publications have been produced in the course of this thesis:

Journals

[J1] S. Lakshminaryana, M. Assaad and M. Debbah, "Energy Efficient Cross-Layer Design in MIMO Multi-cell Systems," **to be submitted**

[J2] S. Lakshminaryana, J. Hoydis, M. Debbah and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," submitted to *IEEE Transactions on Signal Processing,* 2012

[J3] S. Lakshminarayana, M. Assaad and M. Debbah, "H^{∞} Control Based Scheduler for the Deployment of Small Cell Networks," to appear in *Elsevier Performance Evaluation Journal*, Special Issue of selected papers from WiOpt 2011 (accepted for publication)

Conferences

[C1] S. Lakshminaryana, M. Assaad and M. Debbah, "Energy Efficient Cross-Layer Design in MIMO Multi-cell Systems," **to be submitted** International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13), 2013

[C2] S. Lakshminarayana, B. Li, M. Assaad, A. Eryilmaz and M. Debbah, "A Fast-CSMA Based Distributed Scheduling Algorithm under SINR Model" *IEEE International Symposium on Information Theory (ISIT'12)*, Cambrigde, MA, USA, 2012

[C3] S. Lakshminarayana, M. Debbah and M. Assaad, "Asymptotic Analysis of Downlink Multi-cell Systems with Partial CSIT", *IEEE International Symposium on Information Theory (ISIT'11)* , St-Petersburg, Russia, 2011

[C4] S. Lakshminarayana, M. Assaad and M. Debbah, "H^{∞} Control Based Scheduler for the Deployment of Small Cell Networks," *9th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'11)* , Princeton, USA, 2011

[C5] S. Lakshminaryana, J. Hoydis, M. Debbah and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC'10)*, Istanbul, Turkey, 2010

Chapter 2

Theory

In this chapter, we provide some necessary background which will be of use in developing the algorithms in this thesis. Specifically, we provide results from the fields of probability and RMT, stochastic control theory and stochastic network optimization.

2.1 Useful Results from Probability and Matrix Analysis

We first present some basic results related to matrices.

Lemma 1. *([28]) For two matrices* $\mathbf{X}$ *and* $\mathbf{Y}$*, if* $0 \leq \mathbf{X} \leq \mathbf{Y}$[1] *then* $\rho(\mathbf{X}) \leq \rho(\mathbf{Y})$.

Theorem 1. *(Rayleigh-Ritz Theorem [29]) Let* $\mathbf{A} \in \mathbb{C}^{N \times N}$ *be a Hermitian matrix* $\lambda^{\max}(\mathbf{A})$ *be the maximum eigenvalue, then,*

$$\begin{aligned}\lambda^{\max}(\mathbf{A}) &= \max_{\mathbf{x} \in \mathbb{C}^N \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} \\ &= \max_{\mathbf{x} \in \mathbb{C}^N, ||\mathbf{x}||_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}\end{aligned}$$

Lemma 2. *([30], (2.2)) Let* $\mathbf{A}$ *be a Hermitian invertible matrix of size* $N \times N$*, then for any vector* $\mathbf{x} \in \mathbb{C}^N$ *and scalar* $\tau \in \mathbb{C}$ *for which* $\mathbf{A} + \tau \mathbf{x}\mathbf{x}^H$ *is invertible,*

$$\mathbf{x}^H (\mathbf{A} + \tau \mathbf{x}\mathbf{x}^H)^{-1} = \frac{\mathbf{x}^H \mathbf{A}^{-1}}{1 + \tau \mathbf{x}^H \mathbf{A}^{-1} \mathbf{x}}.$$

Next, we present some standard definitions and results from probability theory.

[1] Recall that for matrices, $\mathbf{X} \leq \mathbf{Y}$ stands for element wise inequality.

Lemma 3. *(Markov's inequality, [31], (5.31)) Let X be a nonnegative random variable and $\epsilon > 0$. Then,*

$$P(X \geq \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}[X].$$

In particular, for an arbitrary random variable X and some integer k,

$$P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^k}\mathbb{E}[|X|^k].$$

Lemma 4. *(Holder's inequality, [31],(5.35)) For X, Y two arbitrary random variables and $p, q > 1$, satisfying $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\mathbb{E}\left[|XY|\right] \leq \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \left(\mathbb{E}\left[|Y|^q\right]\right)^{1/q}.$$

In particular, for two sets $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ of complex numbers,

$$\sum_i |x_i y_i| \leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |x_i|^q\right)^{1/q}.$$

We now deal with infinite sequences $X_1(\omega), X_2(\omega), \ldots$ of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1. *(Almost sure convergence). The sequence of random variables $(X_n)_{n\geq 1}$ converges almost surely to X, if*

$$P\left(\limsup_{n\to\infty} |X_n - X| = 0\right) = 1.$$

This is denoted by $X_n \xrightarrow{a.s.} X$.

In order to prove the almost sure convergence of a sequence of random variables $(X_n)_{n\geq 1}$ to some constant X, one often relies on the combination of Markov's inequality and the Borel-Cantelli lemma which we state in the following.

Lemma 5. *Borel Cantelli Lemma ([31], Theorem 4.3): Let $(A_n)_{n\geq 1}$ be a sequence of sets, $A_n \in \mathcal{F}$ for some probability space $(\Omega, \mathcal{F}, P)$. If $\sum_n P(A_n) \leq \infty$, then $P(\limsup_n A_n) = 0$.*

Lemma 6. *([32], Lemma 2.6) Let $(\mathbf{A}_N)_{N\geq 1}$,$\mathbf{A}_N \in \mathbb{C}^{N\times N}$ be a sequence of matrices and $(\mathbf{x}_N)_{N\geq 1}$, $\mathbf{x}_N \sim \mathcal{CN}(0, \frac{1}{N}\mathbf{I}_N) \in \mathbb{C}^N$, be a sequence of random vectors independent of $(\mathbf{A}_N)_{N\geq 1}$. Consider $m \geq 2$. Then there exists a constant C_m independent of N and $\mathbf{A}$ such that*

$$\mathbb{E}\left[\left|\mathbf{x}_N^H \mathbf{A}_N \mathbf{x}_N - \frac{1}{N} tr(\mathbf{A}_N)\right|^m\right] \leq \frac{C_m}{N^{m/2}} ||\mathbf{A}_N||^m.$$

This implies by the Markov inequality and the Borel Cantelli lemma for $||\mathbf{A}_N|| < \infty$, *that*

$$\mathbf{x}_N^H \mathbf{A}_N \mathbf{x}_N - \frac{1}{N} tr(\mathbf{A}_N) \xrightarrow[N\to\infty]{a.s.} 0.$$

Lemma 7. *([30], Lemma 2.6) Let* $z \in \mathbb{C}^+$ *with* $v = \mathrm{Im}(z)$ *and* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $N \times N$ *matrices with* $\mathbf{B}$ *being Hermitian,* $\tau \in \mathbb{R}$, *and* $\mathbf{q} \in \mathbb{C}^N$, *then*

$$|tr\big((\mathbf{B} - z\mathbf{I})^{-1} - (\mathbf{B} + \tau \mathbf{q}\mathbf{q}^H - z\mathbf{I})^{-1})\mathbf{A}\big)| \leq \frac{||\mathbf{A}||}{v}.$$

Lemma 8. *([33], Lemma 1) Let* $(a_n)_{n\geq 1}$, $(\bar{a}_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$ *and* $(\bar{b}_n)_{n\geq 1}$ *denote four infinite sequences of complex random variables. If* $a_n \asymp \bar{a}_n$ *and* $b_n \asymp \bar{b}_n$. *Then,*

- *If* $|a_n|, |\bar{b}_n|$ *and/or* $|\bar{a}_n|, |b_n|$ *are bounded almost surely, then,*

$$a_n b_n \asymp \bar{a}_n \bar{b}_n.$$

- *If* $|a_n|, |\bar{b}_n|^{-1}$ *and/or* $|\bar{a}_n|, |b_n|^{-1}$ *are bounded almost surely, then,*

$$\frac{a_n}{b_n} \asymp \frac{\bar{a}_n}{\bar{b}_n}.$$

2.2 Introduction to Random Matrix Theory

We will now provide a brief introduction to RMT.

Definition 2. *For a Hermitian* $N \times N$ *matrix* $\mathbf{X}$, *we denote the empirical spectral distribution (e.s.d.) by* $F^{\mathbf{X}}$, *which is defined for* $x \in \mathbb{R}$ *as*

$$F^{\mathbf{X}}(x) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{\lambda_j \leq x}(x)$$

where $\{\lambda_1, \ldots, \lambda_N\}$ *are the eigenvalues of* $\mathbf{X}$ *and* $\mathbf{1}_{\lambda_j \leq x}$ *is then indicator function whose value is equal to* 1 *if the eigenvalue* λ_j *is less than* x, 0 *otherwise.*

The relevant aspect of large $N \times N$ Hermitian matrices is that their random e.s.d. $F^{\mathbf{X}}$ converges, as $N \to \infty$ towards a nonrandom distribution F,

$$F^{\mathbf{X}} - F \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

This function F, if it exists, will be called the limit spectral distribution (l.s.d.) of $\mathbf{X}$.

In general, it is tedious to work directly with the distribution functions of the eigenvalues of random matrices. Infact, for many of the matrix models, it is difficult to prove the convergence of the e.s.d. to a limiting distribution. In order to overcome this, we now introduce the concept of Stieltjes transform which is a powerful tool in the analysis of random matrices.

Definition 3. *We denote the Stieltjes transform of the e.s.d* $F^{\mathbf{X}}$ *of the matrix* $\mathbf{X}$ *by* $m_{\mathbf{X}}(z)$ *which is defined as*

$$m_{F^{\mathbf{X}}}(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} d\mathbf{F}^{\mathbf{X}}(\lambda).$$

The Stieltjes transform based approach greatly simplifies the study of large dimensional random matrices. The intuition is the following. For a Hermitian matrix $\mathbf{X}$,

$$\begin{aligned} m_{F^{\mathbf{X}}}(z) &= \int_{\mathbb{R}} \frac{1}{\lambda - z} d\mathbf{F}^{\mathbf{X}}(\lambda) \\ &= \frac{1}{N} \mathrm{tr}(\mathbf{\Lambda} - z\mathbf{I}_N)^{-1} \\ &= \frac{1}{N} \mathrm{tr}(\mathbf{X} - z\mathbf{I}_N)^{-1} \end{aligned}$$

where $\mathbf{\Lambda}$ is the diagonal matrix consisting of eigen values of the matrix $\mathbf{X}$. Working with Stieltjes transform simplifies to working with the matrix $(\mathbf{X} - z\mathbf{I}_N)^{-1}$. From matrix inversion lemmas and several fundamental matrix identities, it is then rather simple to derive limits of traces $\frac{1}{N}\mathrm{tr}(\mathbf{X} - z\mathbf{I}_N)^{-1}$. For many matrix models, it is easier to show that the Stieltjes transform $m_{F^{\mathbf{X}}}$ of the e.s.d. of the matrix $\mathbf{X}$ converges almost surely to a function m, which is itself the Stieltjes transform of a distribution function F, i.e.,

$$m_{F^{\mathbf{X}}} - m \xrightarrow[N\to\infty]{\text{a.s.}} 0.$$

To specify the exact link between the distribution functions and their Stieltjes transform, we state the following theorem.

Theorem 2. *Let* $\{F_N\}$ *be a set of bounded real functions such that* $\lim_{x\to-\infty} F_N(x) = 0$. *Then for all* $z \in \mathbb{C}^+$,

$$\lim_{N\to\infty} m_{F_N}(z) = m_F(z)$$

if and only if there exists F *such that* $\lim_{x\to-\infty} F(x) = 0$ *and* $|F_N(x) - F(x)| \to 0$ *for all* $x \in \mathbb{R}$.

Theorem 2 implies that the convergence of the Stieltjes transform $m_{F_N}(z)$ to $m_F(z)$ ensures the convergence of the distribution functions and vice versa. This justifies the use of Stieltjes transform in the analysis of random matrices. For the sake of brevity, we will call $m_{F^{\mathbf{X}}}(z)$ the Stieltjes transform of $\mathbf{X}$, rather than the Stieltjes transform of $F^{\mathbf{X}}$.

For many applications of practical interest, one can directly work with Stieltjes transform rather than working with the spectral distributions. As an example, consider the SINR at the output of a MMSE receiver which is given as

$$\gamma = \mathbf{h}(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I})^{-1}\mathbf{h}$$

where $\mathbf{h} \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(0, \mathbf{I}_N)$ and the matrix $\mathbf{H} \in \mathbb{C}^{N\times N}$ whose elements are independent of $\mathbf{h}$. From the result of Lemma 6, it follows that

$$\mathbf{h}(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I})^{-1}\mathbf{h} \xrightarrow[N\to\infty]{\text{a.s.}} \frac{1}{N}\text{tr}((\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I})^{-1})$$

where the term on the right hand side is by definition the Stieltjes transform of the matrix $\mathbf{H}\mathbf{H}^H$. Therefore, in many cases it suffices to find the asymptotic equivalent of the Stieltjes transform rather finding the l.s.d.

We will finally state the following theorem which will be of use in this thesis.

Theorem 3. *([30]) Consider the matrix* $\mathbf{B}_{N_t} = \mathbf{X}_{N_t}\mathbf{T}_{N_t}\mathbf{X}_{N_t}^H$, *where* $\mathbf{X}_{N_t} = \frac{1}{\sqrt{N_t}}\mathbf{Y}_{N_t} \in \mathbb{C}^{N_t\times NK}$ *with entries* $\mathbf{Y}_{N_t}(p,q) \sim \mathcal{CN}(0,1)$, *and the matrix* $\mathbf{T}_{N_t}$ *a non random diagonal matrix given by* $\mathbf{T}_{N_t} = diag(t_1, \ldots, t_{NK}) \in \mathbb{R}^{NK\times NK}$. *Let* $m_{\mathbf{B}_{N_t}}(z) = \frac{1}{N_t}tr(\mathbf{B}_{N_t} + z\mathbf{I})^{-1},\ z > 0,$ *Then,*

$$m_{\mathbf{B}_{N_t}}(z) - \bar{m}(z) \xrightarrow[N_t,K\to\infty]{a.s.} 0$$

where $\bar{m}(z)$ *can be evaluated as the unique solution to the fixed point equation*

$$\bar{m}(z) = \left(\frac{1}{N_t}\sum_{i=1}^{NK}\frac{t_i}{1+t_i\bar{m}(z)} + z\right)^{-1}.$$

Under the same assumptions as Theorem 3, let $m'_{\mathbf{B}_{N_t}}(z) = \frac{1}{N_t}\text{tr}(\mathbf{B}_{N_t}+z\mathbf{I})^{-2}$. Then,

$$m'_{\mathbf{B}_{N_t}}(z) - \bar{m}'(z) \xrightarrow[N_t,K\to\infty]{\text{a.s.}} 0$$

where $\bar{m}'(z)$ can be evaluated as

$$\bar{m}'(z) = \frac{\bar{m}^2(z)}{1 - \frac{\bar{m}^2(z)}{N_t}\sum_{i=1}^{NK}\frac{t_i^2}{(1+t_i\bar{m}(z))^2}}.$$

Interestingly, the calculation of $\bar{m}(z)$ and $\bar{m}'(z)$ depends on only the second order statistics of the random matrix $\mathbf{X}$. At this point, we remark that all the above results also hold for a more general non-Gaussian vectors/ matrices satisfying some moment conditions [30].

2.3 Overview of Results from Stochastic Control Theory

In this section, we provide a brief description of two control theoretic algorithms which will be used in this thesis namely the H^∞ control and the linear quadratic Gaussian (LQG) control algorithm.

Let us consider the state space equation

$$\mathbf{x}[t+1] = \mathbf{A}[t]\mathbf{x}[t] + \mathbf{B}[t]\mathbf{u}[t] + \mathbf{D}[t]\mathbf{w}[t] \tag{2.1}$$

where the state vector $\mathbf{x}[t]$, control vector $\mathbf{u}[t]$ and noise vector $\mathbf{w}[t]$ take values respectively in $\mathbb{R}^n$, $\mathbb{R}^p$ and $\mathbb{R}^m$. The objective is to design a controller that minimizes the following cost function:

$$L(\mathbf{u}, \mathbf{w}) = |\mathbf{x}[T]|^2 + \sum_{t=1}^{T-1} \left[|\mathbf{x}[t]|^2 + |\mathbf{u}[t]|^2 \right]. \tag{2.2}$$

Depending on the nature of the noise process $\mathbf{w}[t]$, we describe two control theoretic algorithms. In the first case, we consider $\mathbf{w}[t]$ to be a Gaussian process, in which case the optimal solution is given by the LQG control.

2.3.1 LQG Control

We asume that the sequence $\mathbf{w}[t]$ is Gaussian distributed with zero mean and covariance matrix given by

$$\mathbb{E}\left[(\mathbf{w}[t])^T \mathbf{w}[s] \right] = \mathbf{R}_1 \boldsymbol{\delta}[t, s]$$

where $\mathbf{R}_1$ is the covariance matrix and $\boldsymbol{\delta}[t, s]$ is the Kronecker delta function given by

$$\boldsymbol{\delta}[t, s] = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{else.} \end{cases} \tag{2.3}$$

With the knowledge of the statistics of the noise process, the problem then is to minimize the expected value of the cost function, $\mathbb{E}\left[L(\mathbf{u}, \mathbf{w}) \right]$. We assume that the controller makes perfect observation of the state variable $\mathbf{x}[t]$. The LQG control with perfect state observations is summarized as follows [34]. Starting with the initial state $\mathbf{x}[0] = 0$, solve the Ricatti equations

$$\begin{aligned} \mathbf{M}[t] &= (\mathbf{A}[t])^T \mathbf{M}[t+1] \mathbf{A}[t] \\ &\quad - (\mathbf{A}[t])^T \mathbf{M}[t+1] \mathbf{B}[t] \Big((\mathbf{B}[t])^T \mathbf{M}[t+1] \mathbf{B}[t] + \mathbf{I} \Big)^{-1} (\mathbf{B}[t])^T \mathbf{M}[t+1] \mathbf{A}[t] + \mathbf{Q}[t]; \\ \mathbf{M}[T] &= \mathbf{Q}_f. \end{aligned}$$

The optimal control strategy which is a linear control law in the state variable is

$$\mathbf{u}[t] = -\mathbf{L}[t]\mathbf{x}[t]$$

where the matrix $\mathbf{L}[t]$ is given by

$$\mathbf{L}[t] = \Big((\mathbf{B}[t])^T \mathbf{M}[t+1] \mathbf{B}[t] + \mathbf{I} \Big)^{-1} (\mathbf{B}[t])^T \mathbf{M}[t+1] \mathbf{A}[t].$$

2.3.2 H^∞ Control

Next, we consider the case when the noise is unpredictable and the controller does not have knowledge of the statistical distribution of the noise process. The problem to solve is a linear control problem with quadratic cost and unpredictable noise. In this case, we only make the assumption that the noise process $\mathbf{w}$ is square integrable (of finite energy). When the noise is unpredictable (we don't have information on the statistical distribution of the noise), an efficient way to solve the aforementioned problem is to use the H^∞ control technique [35].

Note that the cost L is a quadratic function of the noise. When the controller has no information about the noise process, there is no way to prevent the cost from becoming large as the L_2 norm of the noise increases. Instead one can think of controlling the rate of growth, i.e. to minimize the ratio $\frac{|\mathbf{x}[t]|^2}{|\mathbf{w}[t]|^2}$, $\forall t$. In order to accomplish this using the H^∞ technique, the quadratic cost to be minimized is reformulated as follows:

$$L_\gamma(\mathbf{u}, \mathbf{w}) = |\mathbf{x}[T]|^2 + \sum_{t=1}^{T-1} \left[|\mathbf{x}[t]|^2 + |\mathbf{u}[t]|^2 - \gamma^2 |\mathbf{w}[t]|^2\right] \tag{2.4}$$

where, γ^2 is the level of attenuation (one can refer to [35] for the exact role of γ in H^∞ formulation). In the absence of statistical distribution of noise, the controller design is a minimax optimization problem, where the cost function L_γ is minimized over maximum of the unknown disturbance. This problem is a two player zero sum game. The cost function L_γ maximized by Player 1 (noise) and minimized by Player 2 (controller) using vectors $\mathbf{w}[t]$ and $\mathbf{u}[t]$ respectively. This implies that the controller design attempts to minimize the cost function assuming the worst case noise (and hence robust controller). One can refer to [35] for more general classes of discrete-time zero-sum games, with various information patterns, where sufficient conditions for the existence of a saddle point are provided when the state information pattern is perfect and imperfect. In particular, in our work, we consider the H^∞ controller with closed loop perfect state information pattern which will be stated in the following proposition:

Proposition 1. *([35], Theorem 3.2) There exists a unique feedback saddle point solution if*

$$\gamma^2 \mathbf{I} - (\mathbf{D}[t])^T \mathbf{M}[t+1] \mathbf{D}[t] > 0, \qquad t \in [1, T],$$

where the sequence of nonnegative definite matrices $\mathbf{M}[t+1], t \in [1, T]$, *are*

generated according to

$$\mathbf{M}[t] = \mathbf{Q}[t] + (\mathbf{A}[t])^T \mathbf{M}[t+1] \mathbf{\Lambda}^{-1}[t] \mathbf{A}[t] \qquad \mathbf{M}[T+1] = \mathbf{Q}_f \tag{2.5}$$

where,

$$\mathbf{\Lambda}[t] = \mathbf{I} + \left(\mathbf{B}[t](\mathbf{B}[t])^T - \gamma^{-2} \mathbf{D}[t] \mathbf{D}^T[t] \right) \mathbf{M}[t]. \tag{2.6}$$

Under the conditions above, there exists a unique feedback saddle point policy given by

$$\mathbf{u}^*[t] = -(\mathbf{B}[t])^T \mathbf{M}[t+1] \mathbf{\Lambda}^{-1}[t] \mathbf{A}[t] \mathbf{x}[t]$$
$$\mathbf{w}^*[t] = \gamma^{-2} (\mathbf{D}[t])^T \mathbf{M}[t+1] \mathbf{\Lambda}^{-1}[t] \mathbf{A}[t] \mathbf{x}[t]$$

with the corresponding unique state trajectory generated by the difference equation

$$\mathbf{x}^*[t+1] = (\mathbf{\Lambda}[t])^{-1} \mathbf{A}[t] \mathbf{x}^*[t], \qquad \mathbf{x}^*[1] = \mathbf{x}[1].$$

2.4 Overview of results from Stochastic Network Optimization

We now explain some basic concepts from the field of stochastic network optimization which are mainly based on the idea of Lyapunov analysis. We first introduce the basic notion of stability which will be used in this thesis. Let $Q[t]$ be a real valued stochastic process that evolves in discrete time over slots $t \in \{0, 1, 2, \dots, \}$ according to some probability law. The discrete time process $Q[t]$ is strongly stable if:

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[|Q[t]|\right] < \infty.$$

We now proceed to consider a set of queues which are served over a wireless medium. Let us denote the set of N queues whose queue length during time t is denoted by $Q_i[t], i = 1, \dots, N$. We can represent the evolution of the queue in the following manner.

$$Q_i[t+1] = \max(Q_i[t] - \mu_i[t], 0) + A_i[t]$$

where $\mu_i[t]$ is the (potentially random) amount of backlog that is transmitted out of queue during the time t and $A_i[t]$ denotes is the (potentially random) amount which arrives into the queue. The departure of packets from the queues in turn depends upon the current network topology and the control action implemented.

The control action could be in the form of scheduling, channel allocation (as in orthogonal frequency division multiplexing (OFDM) based systems), power control or the concept of beamforming (with multiple antennas at the transmitter and receiver). For example, the number of packets departing the queues could be a function of the signal to interference noise ratio (which in turn depends on the link rates and the control policy implemented).

We now present some basic definitions and theorems.

- **Strong Stability** A system of queues is said to be stable if all individual queues of the network are strongly stable.
- **Stability Region** The stability region of a system is the set of all arrival rate vectors for which the system can be stabilized by some control policy. Every arrival rate vector belonging to the stability region is stabilizable, since there exists a control policy under which the system is stable for that arrival rate vector. Moreover, a rate vector outside the stability region is not stabilizable, since all control policies would lead to unbounded queues in the system for that arrival rate vector.
- **Throughput Optimal** A control policy is said to be throughput-optimal if it stabilizes the system for all arrival rate vectors that are strictly within the stability region. In other words, a throughput optimal control policy can stabilize all arrival rate vectors that belong to the interior of the stability region.

2.4.1 Concept of Lyapunov Stability

The concept of Lyapunov stability has been one of the most prominent tool to prove the stability of queuing systems. The idea is to define a non-negative function, called a Lyapunov function, as a scalar measure of the aggregate congestion of all queues in the system. The control decisions are then evaluated in terms of how they affect the change in the Lyapunov function from one slot to the next. We denote the Lyapunov function by the notation $V(\mathbf{Q}[t])$, where $\mathbf{Q}[t]$ is the collection of all the queues in the system. The Lyapunov function $V : \mathbb{R}^N \rightarrow \mathbb{R}$ must have the following properties.

- It should be continuously differentiable function.
- $V(\mathbf{x}) > 0,\ \forall \mathbf{x} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$
- $V(\mathbf{0}) = 0$.

We also define the Lyapunov drift denoted by $\Delta(\mathbf{Q}[t])$ which is given as

$$\Delta(Q[t]) = \mathbb{E}\left[V(\mathbf{Q}[t+1]) - V(\mathbf{Q}[t])\Big|\mathbf{Q}[t]\right], \tag{2.7}$$

where the expectation is with respect to the random channel states and the (possibly random) control actions made in reaction to these channel states given the current queue length $\mathbf{Q}[t]$. The Lyapunov drift represents the expected change in the Lyapunov function from one time slot to the other. The concept of Lyapunov drift can be used to design online control strategies which can ensure queue-length stability. We now state the basic theorem of Lyapunov stability which is as follows.

Theorem 4. *(Lyapunov Stability [13]) If there exist constants $B > 0$, $\epsilon > 0$, and $V > 0$ such that for all time slots t we have,*

$$\mathbb{E}\left[V(\mathbf{Q}[t+1]) - V(\mathbf{Q}[t])\Big|\mathbf{Q}[t]\right] \leq B - \epsilon\sum_{i=1}^{N} Q_i[t],$$

then, the network is strongly stable and we have

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{N}\mathbb{E}\left[Q_i[t]\right] \leq \frac{B}{\epsilon}.$$

The crux of stochastic network optimization lies in designing online control algorithms which ensure the condition provided in Theorem 4. It then naturally follows that the queue-length process which evolves such that the Lyapunov drift satisfies the above stated condition ensures strong stability.

In addition to designing algorithms that ensure queuing stability, the concept of Lyapunov drift can be used to optimize an additional performance metric of interest. For example, of all the control actions that maintain the stability of a system of queues, one might be interested in choosing a policy that yields the least expenditure in terms of the a energy consumption. Let us denote this additional performance metric of interest by the notation $\zeta[t]$. For the sake of explanation, we assume $\zeta[t]$ to be a non-negative penalty function which we are trying to minimize. We denote its time average expectation by

$$\bar{\zeta} = \limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\zeta[t]\right].$$

Let us define ζ^* to be the infimum time average penalty $\bar{\zeta}$ incurred over all possible sequences on control actions. The following result allows us to consider the result of queuing stability and performance optimization using a single drift analysis.

Theorem 5. *(Lyapunov Optimization [13]) If there exist constants $B > 0$, $\epsilon > 0$ such that for all time slots t we have,*

$$\Delta(\mathbf{Q}[t]) + V\mathbb{E}\left[\zeta[t]\middle|\mathbf{Q}(t)\right] \leq B - \epsilon\sum_{i=1}^{N} Q_i[t] + V\zeta^*,$$

then the system is stable and the time average backlog satisfies:

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{N}\mathbb{E}\left[Q_i[t]\right] \leq \frac{B + V\zeta^*}{\epsilon}$$

and

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\zeta[t]\right] \leq \zeta^* + B/V.$$

The parameter V represents the trade-off between how close one can get to the optimal value of the parameter of interest and the average queue-length. Increasing V can result in getting arbitrarily close to the optimal value at the expense of increasing the average backlog (and hence the average delay). Developing a control policy which ensures the condition of Theorem 5 ensures not only the queue length stability but also optimizes the performance metric of interest.

With this background, we now proceed to pose cross-layer design problems in MIMO multi-cell scenario and provide solutions.

Chapter 3

Beamforming Design and Traffic Flow Control

Abstract: *In this chapter, we address the joint problem of traffic flow control and interference management in a MIMO multi-cell network setup considered in Figure 3.1. Future cellular networks are envisioned to be composed of small cell networks (SCNs) [3]. The BSs of the SCNs are low power devices which have limited sized buffers to store the packets of the UTs they serve. In this scenario, traffic flow from the network to the buffers present at the BS must be regulated in such a way that the queue-length (of the buffer) at BS remains as close as possible to the target queue-length (usually the size of the buffer). This task is normally accomplished with the help of a flow controller which regulates the flow of packets from the network to the queues present at the BSs. Typically, the flow controller is connected to a large number of SCN BSs and hence, cannot have the CSI of the wireless links between the BSs and the UTs. Therefore, the flow controller design must be oblivious to the CSI of the wireless links. On the other hand, the BSs have to transmit the packets arriving into the queues in an efficient manner over the wireless channel to the UTs in an interference limited environment. Centralized interference management techniques call for lot of overhead in terms of information exchange. In this chapter, we decouple the joint problem of traffic flow control and interference management into two separate parts.*

For the problem of interference management over the wireless channel, we formulate a multi-cell beamforming algorithm which requires limited amount of information exchange between the BSs. The design objective is to minimize the total transmit power across all the BSs subject to satisfying the user SINR

constraints. With our algorithm, the BSs need to exchange parameters which can be computed solely based on the channel statistics rather than the instantaneous CSI. We make use of tools from RMT to formulate the distributed algorithm. For the traffic flow control problem, we propose a H^∞ based flow controller which regulates the traffic flow into the queues present at the BS while being oblivious to the CSI of the wireless link between the BSs and the UTs. Our simulation results show that the reduced over head beamforming algorithm coupled with the H^∞ based flow control stabilizes the queues at the BS and limits the variation of the queue-length around the target buffer size.

3.1 Introduction

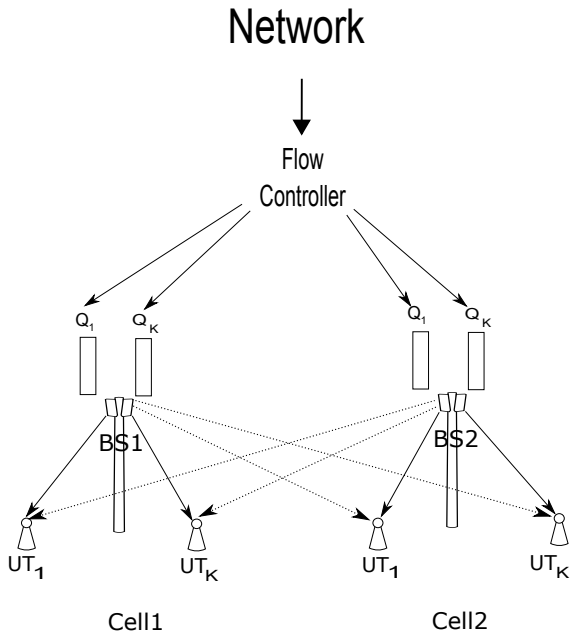

Figure 3.1: System Setup

The chapter is organized as follows. We first provide the system model considered in this chapter and explain the joint problem of traffic flow control and interference management. We explain explicitly how these two problems are inter-related. Later on, we decouple the two problems and describe the solutions to the two parts in detail. In the first part, we deal exclusively with the problem of interference management. Specifically, we present a decentralized multi-cell beamforming technique formulated with the help of tools from RMT.

We provide theoretical analysis for the multi-cell beamforming algorithm design. In the second part of this chapter, we provide the motivation and formulate a H^∞ based flow controller which regulates the traffic arrival into the queues. We also provide some simulation results for both the parts.

3.2 System Model

The system model is shown in Figure 3.1. It can be broken down into two parts. The frontend consisting of the BSs and the UTs and the backend consisting of the flow controller.

We focus on the frontend first. We consider a MIMO multi-cell scenario consisting of N cells and K UTs per cell. The UTs in each cell are served by their respective BSs which are equipped with N_t antennas. The UTs have a single antenna each. The notation $\mathrm{UT}_{j,k}$ denotes the k-th UT present in the j-th cell. The BS of each cell serves only the UTs present in its cell. We consider a discrete-time block-fading channel model where the channel remains constant for a given coherence interval and then changes independently from one block to the other. We index the coherence intervals by t and address them as time slots. We assume that the elements of the channel vector from BS i to $\mathrm{UT}_{j,k}$ during the time slot t are given by $\mathbf{h}_{i,j,k}[t] = \sqrt{\sigma_{i,j,k}/N_t}\mathbf{x}_{i,j,k}[t]$, where the vector $\mathbf{x}_{i,j,k}[t]$ consists of i.i.d. random variables with zero mean and variance 1. $\sigma_{i,j,k}$ are real numbers that depend upon the path loss between BS i and $\mathrm{UT}_{j,k}$. Additionally, we make the assumption that the random vector $\mathbf{x}_{i,j,k}[t]$ has bounded L2 norm [1].

We let $\mathbf{w}_{i,j}[t] \in \mathbb{C}^{N_t}$ denote the transmit downlink beamforming vector for the $\mathrm{UT}_{i,j}$ during time slot t. The received signal $y_{i,j}[t] \in \mathbb{C}$ for the $\mathrm{UT}_{i,j}$ during the time slot t is given by

$$y_{i,j}[t] = \sum_{l=1}^{K} \mathbf{h}_{i,i,j}^{H}[t]\mathbf{w}_{i,l}[t]x_{i,l}[t] + \sum_{\substack{m=1 \\ m\neq i}}^{N}\sum_{n=1}^{K} \mathbf{h}_{m,i,j}^{H}[t]\mathbf{w}_{m,n}[t]x_{m,n}[t] + z_{i,j}[t] \quad (3.1)$$

where $x_{i,j}[t] \in \mathbb{C}$ represents the information signal for the $\mathrm{UT}_{i,j}$ during the time slot t and $z_{i,j}[t] \sim \mathcal{CN}(0, N_0)$ is the corresponding additive white Gaussian complex noise. Let $\Gamma_{i,j}[t]$ denote the achieved SINR in the downlink for $\mathrm{UT}_{i,j}$ and $\gamma_{i,j}[t]$ the corresponding target SINR during the time slot t. The achieved

[1] The bounded L2 norm is an important assumption for the modeling of the H^∞ controller which will be explained later.

SINR in the downlink is given by the following expression

$$\Gamma_{i,j}[t] = \frac{|\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2}{\sum_{l\neq j}|\mathbf{w}_{i,l}^H[t]\mathbf{h}_{i,i,j}[t]|^2 + \sum_{m\neq i,n}|\mathbf{w}_{m,n}^H[t]\mathbf{h}_{m,i,j}[t]|^2 + \sigma^2}. \tag{3.2}$$

The numerator term is the useful signal. The denominator terms represent the intra-cell interference, inter-cell interference and the thermal noise (in order as they appear in the denominator).

The backend part consists of the flow controller in between the network and the queues present at the BS. The task of the flow controller is to regulate the traffic arriving from the network into the queues present at the BS. We denote the queue-length of the $\mathrm{UT}_{i,j}$ during the time slot t by $Q_{i,j}[t]$. The number of packets arriving into the queue during the time slot t is denoted by $A_{i,j}[t]$. The arrival process is regulated by the flow controller. We denote the number of packets transmitted during the time slot t by the notation $\mu_{i,j}[t]$. The number of packets departing the queue depends on the achieved SINR in the downlink given by the relation [2]

$$\mu_{i,j}[t] = B\log(1+\Gamma_{i,j}[t]). \tag{3.3}$$

where B is the total number of channel uses available during the coherence interval T_c. The queue-length evolves according to the following equation,

$$Q_{i,j}[t+1] = Q_{i,j}[t] + A_{i,j}[t] - \mu_{i,j}[t] + U_{i,j}[t] \qquad \forall i,j \tag{3.4}$$

where $U_{i,j}[t] = B < \mu_{i,j}[t]$ if $Q_{i,j}[t] \leq \mu_{i,j}[t]$, else $U_{i,j}[t] = 0$. It has the effect of keeping the queue-lengths positive at all time instants. Let us denote the target buffer length at the BS by the notation $\bar{Q}_{i,j}$ (which is also the capacity of the buffer).

The downlink power minimization and flow controller design problem can be cast into the following optimization problem given by

$$\begin{aligned} \min_{\mathbf{w}_{i,j}[t]} \quad & \sum_{i,j}\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t], \qquad \forall t = 1\ldots T \\ s.t. \quad & \Gamma_{i,j}[t] \geq \gamma_{i,j}[t], \qquad t = 1\ldots T, i = 1\ldots N,\ j = 1\ldots K \\ & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} Q_{i,j}[t] = \bar{Q}_{i,j} \end{aligned} \tag{3.5}$$

The objective function of the optimization problem in (3.5) attempts to minimize the total power transmitted in the downlink during each time slot t. The first constraint equation is the target SINR constraint for each UT. The second

[2] We assume that the coherence interval of the channel is long enough for the transmitter to achieve the channel capacity.

constraint equation denotes the flow controller design which tries to maintain the time average of the queue-length as close as possible to the target. We assume that the target SINR constraints $\gamma_{i,j}[t]$ are pre-computed according to the demands of the UTs (we do not optimize over the choice of $\gamma_{i,j}[t]$).

The target queue-lengths (or the buffer capacity) in this work can be viewed as a way to model the fairness criteria between different UTs. The target queue-lengths are a representative of the average number of packets served to each UT. They can be designed to be different for different UTs depending on the fairness objective.

3.3 Algorithm Design

The basic optimization problem in (3.5) is difficult to solve due to the interdependencies of the various parameters involved. In this work, we do not exactly solve (3.5). Instead, in this work, we decouple the problem into two parts. The first part is the flow control problem for the queues of the BS. In this part, the design objective considered is to minimize the variation of the queue-length around the target. The second part is the interference management problem. The decoupled algorithm can be summarized as follows:

- The UTs request a target SINR during each time slot depending on their requirement $\gamma_{i,j}[t]$.
- The BSs then jointly solve the following optimization problem

$$\min_{\mathbf{w}_{i,j}\in\mathbb{C}^{N_t}} \quad \sum_{i,j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \tag{3.6}$$

$$s.t. \quad \frac{|\mathbf{w}_{i,j}^H \mathbf{h}_{i,i,j}[t]|^2}{\sum_{l\neq j} |\mathbf{w}_{i,l}^H \mathbf{h}_{i,i,j}[t]|^2 + \sum_{m\neq i,n} |\mathbf{w}_{m,n}^H \mathbf{h}_{m,i,j}[t]|^2 + \sigma^2} \geq \gamma_{i,j}[t], \qquad \forall i,j$$

  in order to compute the optimal beamforming vectors during each time slot. To this end, we formulate a decentralized multi-cell beamforming algorithm.

- The number of packets transmitted from the queue of $\mathrm{UT}_{i,j}$ during the time slot t depend on the achieved SINR given by the relation (3.3). The achieved SINR in the downlink in turn depends on the channel conditions between the BSs and the UTs during the corresponding time slot.
- Since the flow controller is typically connected to a large number of BSs, it does not have knowledge of the wireless channel conditions between the BSs and the UTs. This implies that it does not have knowledge of

the actual achieved SINR in the downlink (and hence number of packets transmitted out of the queue during the time slot t, $\mu_{i,j}[t]$). Therefore, the flow controller must regulate the number of packets flowing into the queues with the objective of minimizing the fluctuations of the queue-length around the target, while being oblivious to the wireless channel conditions.

In the next part, we exclusively consider the problem of multi-cell beamforming design of (3.6).

3.4 Part I: Interference Management - Multi-cell Beamforming

We first start with the decentralized multi-cell beamforming design algorithm. We provide a brief background of multi-cell beamforming design problem and the past works. Later on, we will describe our decentralized multi-cell beamforming algorithm and provide a theoretical analysis.

3.4.1 Introduction to Multi-cell Beamforming Design

The downlink of a cellular system has been a widely researched topic in the past decade both in the isolated single cell processing scenario [36] and the joint multi-cell processing scenario [37, 38, 39]. It has been well established that significant performance improvements can be obtained if the BSs coordinate by CSI and user data sharing via high-capacity backhaul links (network MIMO) and jointly serve the UTs [37, 40]. However, as the number of antennas on the BS, the number of UTs and the number of coordinating cells grow large, this approach quickly becomes impractical due to the heavy backhaul requirement. Hence, there has been emphasis on developing distributed strategies that exploit only the locally available CSI.

Our focus of the multi-cell beamforming design of this chapter is primarily based on the *coordinated beamforming* approach [41, 42, 43] in which BSs formulate their beamforming vectors taking into consideration the inter-cell interference. However, unlike the case of network MIMO, no exchange of user data information takes place between the BSs. Coordinated beamforming can be considered as a mid-way between network MIMO on the one hand and single cell processing on the other hand. Some of the relevant works based on joint multi-cell beamforming (without user data sharing) include [44, 45, 46, 47] in the multiple input single output interference channel (MISO IC) scenario. Other

relevant references which consider joint beamforming with limited information exchange between the BSs include [48, 49].

Reference [42] provides an optimal algorithm for the multi-cell beamforming problem where each BS is serving an arbitrary number of UTs. Here, the design objective is to minimize the total transmitted power satisfying the SINR constraints of the UTs. In order to compute the optimal beamforming vectors, the BSs must solve an algorithm for every channel realization and then exchange this information between themselves. All this has to be done within a given coherence interval. Such a strategy demands high computational ability and rapid information exchange between the BSs, especially in fast fading scenarios.

In order to overcome the heavy backhaul requirement, we propose in this chapter, an alternative approach to compute multi-cell beamforming vectors. In our algorithm, the BSs compute and exchange parameters at the time scale at which the channel statistics change rather than the instantaneous channel realizations. We use tools from RMT to formulate our algorithm.

Before we proceed, we remark that RMT has been extensively used in analyzing the performance of communication systems. The most attractive feature of RMT results is that it provides compact and elegant expressions which are much easier to analyze than the expressions obtained from finite dimensional analysis. These results often provide good approximations for finite dimensional scenarios. Such results have been exploited in many works. To name a few, the analysis of CDMA systems [50], analysis of linear multi-user receivers [51, 52], sum capacity analysis of MIMO systems under various channel model assumptions [53, 54, 55], optimizing the training time for network MIMO systems [56] and linear precoders [57]. For a detailed survey of results regarding RMT results applied to communication systems, the reader can refer to [58] (Chapter 11). In contrast to the many previous works in which RMT is used as a tool for the performance analysis of the system, in this work we use RMT to optimize the system design parameters (transmit beamforming vectors).

Before we proceed, we would like to place the contributions of the algorithm developed in this part of the chapter in the context of the existing works. The multi-cell beamforming strategy involving exchange of parameters based on channel statistics was first proposed in our related work [59] using tools from RMT. It was shown with the help of simulations that such an algorithm performs well for practical system dimensions. The theoretical analysis regarding the asymptotic optimality of such an algorithm was an open question. The authors in [60] developed multi-cell beamforming algorithms based on the same idea and made arguments about the asymptotic optimality of the same. How-

ever, they consider a two cell Wyner model and symmetric SINR constraints for all the UTs. In this work, we prove the asymptotic optimality of the RMT based distributed beamforming algorithm for a cellular network model with distance based path-loss model. Our main contributions in this part of the chapter are as follows:

- We propose a *reduced overhead* multi-cell beamforming algorithm. In our algorithm, the BSs require only the local CSI (the channels from the given BS to all the UTs present in the system). The BSs compute parameters that depend only on the channel statistics which they exchange between them to compute the beamforming vectors. Additionally, the BSs have to exchange these parameters only at the time scale of channel statistics rather than instantaneous channel realization.
- Using a large system analysis, we provide by closed form expressions a lower bound on the feasible set of target SINR values for which the asymptotic downlink power minimization problem is feasible.
- Further, we prove that this algorithm is asymptotically optimal in the sense that when the dimensions of the system become large, the performance of our algorithm perfectly matches that of the optimal algorithm proposed in [42].

Our results show that the reduced overhead beamforming algorithm provides a good performance for a system with 10 UTs per cell and 10 antennas per BS.

3.4.2 Beamforming Design Algorithm Description

We consider the downlink power minimization problem in (3.5). We take up the beamformer design problem during a given time slot t. We will drop the superscript t in subsequent equations related to the above problem. The optimization problem for the beamforming design is given by

$$\begin{aligned} \min_{\mathbf{w}_{i,j}} \quad & \sum_{i,j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \\ s.t. \quad & \Gamma_{i,j} \geq \gamma_{i,j}, \qquad i = 1 \dots N,\, j = 1 \dots K \end{aligned} \tag{3.7}$$

The downlink sum power minimization problem in (3.7) has been solved in [42] using the second order conic programming (SOCP) framework first developed in [61]. The dual uplink problem corresponding to the optimization in (3.7) is

formulated as

$$\begin{aligned} \max \quad & \sum_{i,j} \lambda_{i,j} N_0 \\ s.t. \quad & \Lambda_{i,j} \geq \gamma_{i,j}, \qquad \forall i,j \end{aligned} \tag{3.8}$$

where the left hand side of the constraint equation represents the uplink SINR given by

$$\Lambda_{i,j} = \frac{\lambda_{i,j} |\hat{\mathbf{w}}_{i,j}^H \mathbf{h}_{i,i,j}|^2}{\sum_{(n,k)\neq(i,j)} \lambda_{n,k} |\hat{\mathbf{w}}_{i,j}^H \mathbf{h}_{i,n,k}|^2 + ||\hat{\mathbf{w}}_{i,j}||_2^2}$$

where $\hat{\mathbf{w}}_{i,j}$ denotes the corresponding uplink receive filter and $\lambda_{i,j}$ represents the dual variable associated with the optimization problem in (3.7). The $\lambda_{i,j}$ can be seen as the dual uplink power.

Algorithm Design: We now provide a brief description of the beamforming algorithm presented in [42]. Before introducing the algorithm, we define the following matrices.

$$\begin{aligned} &\mathbf{H}_{i,n} = [\mathbf{h}_{i,n,1}, \ldots, \mathbf{h}_{i,n,K}] \in \mathbb{C}^{N_t \times K} \\ &\mathbf{H}_i = [\mathbf{H}_{i,1}, \ldots, \mathbf{H}_{i,N}] \in \mathbb{C}^{N_t \times NK} \\ &\boldsymbol{\lambda}_i = [\lambda_{i,1}, \ldots, \lambda_{i,K}] \in \mathbb{C}^{K}, \\ &\boldsymbol{\Lambda} = \operatorname{diag}[\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N] \in \mathbb{C}^{NK \times NK}. \end{aligned}$$

We also define the matrix $\boldsymbol{\Sigma}_i^{\lambda} = \mathbf{H}_i \boldsymbol{\Lambda} \mathbf{H}_i^H \in \mathbb{C}^{N_t \times N_t}$.

[Centralized Algorithm - **CBF**]

Perform the following steps.

- Starting from any initial $\lambda_{i,j}^0 > 0 \; \forall i,j$ the uplink power allocation is given by $\lambda_{i,j} \triangleq \lim_{p\to\infty} \lambda_{i,j}^p$, where

$$\lambda_{i,j}^{p+1} = \frac{1}{(1+\frac{1}{\gamma_{i,j}}) \mathbf{h}_{i,i,j}^H (\boldsymbol{\Sigma}_i^{\lambda^p} + \mathbf{I}_{N_t})^{-1} \mathbf{h}_{i,i,j}} \; \forall i,j \tag{3.9}$$

  where $\boldsymbol{\Sigma}_i^{\lambda^p} = \mathbf{H}_i \boldsymbol{\Lambda}^p \mathbf{H}_i^H$ and $\boldsymbol{\Lambda}^p = \operatorname{diag}[\boldsymbol{\lambda}_1^p, \ldots, \boldsymbol{\lambda}_N^p]$.

- The optimal receive uplink beamforming vectors are given by

$$\hat{\mathbf{w}}_{i,j} = \Big(\sum_{n,k} \lambda_{n,k} \mathbf{h}_{i,n,k} \mathbf{h}_{i,n,k}^H + \mathbf{I} \Big)^{-1} \mathbf{h}_{i,i,j}. \tag{3.10}$$

- The optimal transmit downlink beamforming vectors are given by $\mathbf{w}_{i,j} = \sqrt{\delta_{i,j}} \hat{\mathbf{w}}_{i,j}$, where $\delta_{i,j}$ is given as

$$\boldsymbol{\delta} = \mathbf{F}^{-1} \mathbf{1} N_0. \tag{3.11}$$

Here,

$$\boldsymbol{\delta}_i = [\boldsymbol{\delta}_{i,1}, \ldots, \boldsymbol{\delta}_{i,K}] \in \mathbb{C}^K$$
$$\boldsymbol{\delta} = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_N] \in \mathbb{C}^{NK}$$
$$\mathbf{1} \in [1, \ldots, 1]^T \in \mathbb{R}^{NK}$$

and the elements of the matrix $\mathbf{F} \in \mathbb{C}^{NK \times NK}$ and the submatrix $\mathbf{F}^{i,j} \in \mathbb{C}^{K \times K}$ are given by,

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^{1,1} & \ldots & \mathbf{F}^{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{F}^{N,1} & \ldots & \mathbf{F}^{N,N} \end{pmatrix} \tag{3.12}$$

$$\mathbf{F}^{i,n}_{j,k} \triangleq \begin{cases} \frac{1}{\gamma_{i,j}} |\hat{\mathbf{w}}_{i,j}^H \mathbf{h}_{i,i,j}|^2, & n = i,\ k = j \\ -|\hat{\mathbf{w}}_{i,k}^H \mathbf{h}_{i,i,j}|^2, & n = i,\ k \neq j \\ -|\hat{\mathbf{w}}_{n,k}^H \mathbf{h}_{n,i,j}|^2, & n \neq i. \end{cases} \tag{3.13}$$

As mentioned before, the solution provided in [42] cannot be implemented in a distributed manner. The computation of dual uplink power ($\lambda_{i,j}$) and the scaling factors ($\delta_{i,j}$) requires a central station which has the global CSI knowledge. In what follows, we overcome this problem. We will now formulate our *reduced overhead* beamforming algorithm. We will hereby represent the dual uplink power, the uplink and downlink beamforming vectors of the decentralized algorithm by the notation $\mu_{i,j}, \hat{\mathbf{v}}_{i,j}$ and $\mathbf{v}_{i,j}$, respectively, which are the counterparts of $\lambda_{i,j}, \hat{\mathbf{w}}_{i,j}$ and $\mathbf{w}_{i,j}$ of the CBF algorithm.

[Reduced Overhead Beamforming algorithm - **ROBF**]
Perform the following steps.

- Starting from any initial $\mu_{i,j}^0 > 0\ \forall i, j$ the uplink power allocation is given by $\mu_{i,j} \triangleq \lim_{p\to\infty} \mu_{i,j}^p$, where

$$\mu_{i,j}^{p+1} = \frac{\gamma_{i,j}}{\sigma_{i,i,j} \bar{m}_i^p} \qquad \forall i, j \tag{3.14}$$

  and $\bar{m}_i^p$ is evaluated as $\bar{m}_i^p \triangleq \lim_{q\to\infty} \bar{m}_i^{p,q}$ (initializing with any $\bar{m}_i^{p,0} > 0, \forall i$)

$$\bar{m}_i^{p,q} = \left(\frac{1}{N_t} \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{\sigma_{i,n,k} \mu_{n,k}^p}{1 + \sigma_{i,n,k} \mu_{n,k}^p \bar{m}_i^{p,q-1}} + 1 \right)^{-1}. \tag{3.15}$$

- The optimal receive uplink beamforming vectors are given by

$$\hat{\mathbf{v}}_{i,j} = \Big(\sum_{n,k} \mu_{n,k} \mathbf{h}_{i,n,k} \mathbf{h}_{i,n,k}^H + \mathbf{I} \Big)^{-1} \mathbf{h}_{i,i,j}. \tag{3.16}$$

- The optimal transmit downlink beamforming vectors are given by $\mathbf{v}_{i,j} = \sqrt{\bar{\delta}_{i,j}}\hat{\mathbf{v}}_{i,j}$. The scaling factor $\bar{\delta}_{i,j}$ is given as

$$\bar{\boldsymbol{\delta}} = (\mathbf{I} - \boldsymbol{\Gamma}\boldsymbol{\Delta})^{-1}\boldsymbol{\rho}, \tag{3.17}$$

where

$$\begin{aligned}
&\bar{\boldsymbol{\delta}}_i = [\bar{\delta}_{i,1}, \bar{\delta}_{i,2}, \ldots, \bar{\delta}_{i,K}]^T \in \mathbb{R}^K \\
&\bar{\boldsymbol{\delta}} = [\bar{\boldsymbol{\delta}}_1, \bar{\boldsymbol{\delta}}_2, \ldots, \bar{\boldsymbol{\delta}}_N]^T \in \mathbb{R}^{NK} \\
&\boldsymbol{\gamma}_i = \left[\frac{\gamma_{i,1}}{\sigma_{i,i,1}\bar{G}_{i,i,1}\bar{m}_i^2}, \ldots, \frac{\gamma_{i,K}}{\sigma_{i,i,K}\bar{G}_{i,i,K}\bar{m}_i^2}\right]^T \\
&\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N]^T \\
&\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})
\end{aligned}$$

and the matrix $\boldsymbol{\Delta} \in \mathbb{C}^{NK\times NK}$ is defined as

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}^{1,1} & \ldots & \boldsymbol{\Delta}^{1,N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Delta}^{N,1} & \ldots & \boldsymbol{\Delta}^{N,N} \end{pmatrix} \tag{3.18}$$

where each submatrix $\boldsymbol{\Delta}^{i,j} \in \mathbb{C}^{K\times K}$ is given by

$$\boldsymbol{\Delta}_{j,k}^{i,n} \triangleq \begin{cases} 0, & n = i,\ k = j \\ \frac{1}{N_t}\bar{G}_{i,i,j}\bar{G}_{i,i,k}\bar{m}_i', & n = i,\ k \neq j \\ \frac{1}{N_t}\bar{G}_{n,i,j}\bar{G}_{n,n,k}\bar{m}_n', & n \neq i. \end{cases} \tag{3.19}$$

$\bar{m}_i'$ can be evaluated from $\bar{m}_i$ as

$$\bar{m}_i' = \frac{\bar{m}_i^2}{1 - \frac{1}{N_t}\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{(\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}} \tag{3.20}$$

and the terms

$$\begin{aligned}
\bar{G}_{i,i,j} &= \frac{\sigma_{i,i,j}}{(1+\mu_{i,j}\sigma_{i,i,j}\bar{m}_i)^2} \\
\bar{G}_{i,i,k} &= \frac{\sigma_{i,i,k}}{(1+\mu_{i,k}\sigma_{i,i,k}\bar{m}_i)^2} \\
\bar{G}_{n,i,j} &= \frac{\sigma_{n,i,j}}{(1+\mu_{i,j}\sigma_{n,i,j}\bar{m}_n)^2} \\
\bar{G}_{n,n,k} &= \frac{\sigma_{n,n,k}}{(1+\mu_{n,k}\sigma_{n,n,k}\bar{m}_n)^2}.
\end{aligned} \tag{3.21}$$

The vector $\boldsymbol{\rho}_i = \left[\frac{N_0}{\sigma_{i,i,1}\bar{G}_{i,i,1}\bar{m}_i^2}, \ldots, \frac{N_0}{\sigma_{i,i,K}\bar{G}_{i,i,K}\bar{m}_i^2}\right]^T$ and $\boldsymbol{\rho} = [\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots, \boldsymbol{\rho}_N]^T \in \mathbb{R}^{NK}$.

Remarks:

- We prove the convergence of (3.14) in Section 3.7.1.
- For the details of the computation of $\bar{\boldsymbol{\delta}}$, please refer to Section 3.7.2.

In the rest of the chapter, we will address the above decentralized beamforming algorithm by the acronym ROBF. We will now discuss the practical advantages of the ROBF algorithm over the CBF algorithm.

3.4.3 ROBF Algorithm - Discussion

Notice that the computation of the uplink power allocation in the ROBF algorithm depends only on the second order statistics of the channel matrix and not on the instantaneous CSI. This results in a tremendous reduction of the amount of information to be exchanged between the BSs. In a typical fast fading channel, the channel statistics do not vary rapidly where as the instantaneous channel realizations do. As an example, consider a multi-cell scenario with each BS having 8 antennas serving 4 UTs in each cell and assume that the channel statistics remain constant over 10 different channel realizations. In the CBF algorithm, the BSs have to exchange 320 complex channel coefficients to compute the optimal beamforming vectors during each time slot. However, in the ROBF algorithm the BSs only have to exchange 32 real numbers (the channel statistics) during the 10 slots. The price to pay for the reduction in information exchange between the BSs is that in the ROBF algorithm, the target SINR values are not met perfectly for every channel realization. In fact, for finite values of system dimensions, the achieved SINR in the downlink fluctuates around the target SINR. We will later demonstrate these fluctuations with the help of simulations results in Section 3.4.5. In the next subsection, we will show with rigorous theoretical analysis that the performance of ROBF algorithm perfectly matches the CBF algorithm when the number of antennas per BS and the number of UTs become large.

3.4.4 ROBF Algorithm Analysis

We first start with the analysis of the uplink of the ROBF algorithm. We prove that the achieved SINR in the uplink for the ROBF algorithm converges to the target SINR $\gamma_{i,j}$ when the dimensions of the system grow large. Before we proceed, we make the following observations. Recall that the achieved uplink SINR with arbitrary uplink power allocations $\beta_{i,j}$ (which may or may not depend

on the channel matrix) is given by

$$\Lambda_{i,j}(\beta_{i,j}) = \sigma_{i,i,j}\beta_{i,j}\mathbf{h}_{i,i,j}^H(\mathbf{\Sigma}_i'^{\beta} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j} \tag{3.22}$$

where $\mathbf{\Sigma}_i'^{\beta} = \mathbf{\Sigma}_i^{\beta} - \beta_{i,j}\mathbf{h}_{i,i,j}\mathbf{h}_{i,i,j}^H$ and $\mathbf{\Sigma}_i^{\beta} = \sum_{i,j}\beta_{i,j}\mathbf{h}_{i,i,j}\mathbf{h}_{i,i,j}^H$. Therefore, the achieved SINR in the uplink with power allocations corresponding to the ROBF algorithm is given by $\Lambda_{i,j}(\mu_{i,j}) = \sigma_{i,i,j}\mu_{i,j}\mathbf{h}_{i,i,j}^H(\mathbf{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j}$.

Next, we will show that the uplink SINR with the power allocation from the ROBF algorithm converges almost surely to the target SINR $\gamma_{i,j}$.

Lemma 9. *(Convergence of the Uplink SINR and Uplink Power Allocation)*

$$\Lambda_{i,j}(\mu_{i,j}) \xrightarrow[N_t,K\to\infty]{a.s.} \gamma_{i,j} \qquad \forall i,j. \tag{3.23}$$

The details of the proof can be found in Section 3.7.3.

Next, we state the conditions for feasibility of the downlink power minimization problem.

Closed form expressions for Feasibility Conditions

Note that the solution of the optimization problem using ROBF algorithm consists of two parts: the uplink and the downlink problem. Hence, for the optimization problem to be feasible, both the uplink and the downlink problems should be feasible. First, we will examine the relation between the feasibility of the uplink and the downlink problems.

Lemma 10. *For the ROBF algorithm, the feasibility of the uplink problem implies that the downlink problem is feasible and the matrix* $\mathbf{I}-\mathbf{\Gamma\Delta}$ *is invertible (and thus* (3.17) *has a unique solution).*

The proof can be found in Section 3.7.4, part I.

The above lemma enables us to only look at the feasibility conditions for the uplink. The feasibility of the uplink in a general multi-cell scenario is difficult to be characterized in closed form. Therefore, we consider a new uplink system in which the inter-cell interference path loss coefficients are replaced by $\sigma_{max}(n)$, where $\sigma_{max}(n)$ is the path loss of the channel between the BS of cell i and the edge of the cell which is at a distance n cells from cell i (the neighboring cell would have index 1 and so on - shown in Figure 3.2). Since the distance of any UT inside the cell n is greater than the cell edge distance, the corresponding path-loss $\sigma_{i,n,k} \leq \sigma_{max}(n)$ (recall that the path-loss is inversely proportional to the distance). We call this system as the *modified system*. In what follows,

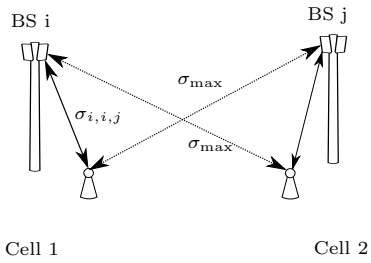

Figure 3.2: Modified system in which the inter-cell interference path loss coefficients are scaled up to σ_{max}.

we will characterize the feasible SINR region for the *modified system*. The feasibility conditions for the uplink of this system will act as a lower bound on the feasibility condition of the original system. Intuitively, this is not hard to see. In the *modified system*, the path losses corresponding to the inter-cell interference links are scaled up to $\sigma_{max}(n)$. Therefore, the *modified system* represents a more interference limited regime as compared to the original system. Hence, any SINR feasible for the *modified system* should be feasible for the original system as well. We will later on make rigorous arguments to prove that the feasibility conditions of the *modified system* are indeed a lower bound on the feasibility conditions of the original system.

Proposition 2. *The ROBF problem is asymptotically feasible if the following holds for* $i = 1, \ldots, N$,

$$\limsup_{N_t, K \to \infty} \left[\frac{1}{N_t} \sum_{k=1}^{K} \frac{\gamma_{i,k}}{1+\gamma_{i,k}} + \frac{1}{N_t} \sum_{\substack{n=1 \\ n \neq i}}^{N} \sum_{k=1}^{K} \frac{\frac{\sigma_{max}(n)}{\sigma_{n,n,k}} \gamma_{n,k}}{1 + \frac{\sigma_{max}(n)}{\sigma_{n,n,k}} \gamma_{n,k}} \right] < 1. \tag{3.24}$$

The proof is provided in Section 3.7.4, part II.

Finally, we will show that the feasibility condition of (3.24) will act as a lower bound on the feasibility conditions of the original system.

Lemma 11. *In the asymptotic limit, any SINR achievable for the* modified system *is achievable by the original system. Hence the achievable SINR region for this system is a lower bound for the original system.*

The details of the proof is provided in Section 3.7.4, part III.

We end this section by providing the feasibility conditions in two special cases which will give us a a good intuition.

Corollary 1: Single Cell Case

For the isolated single cell case, (3.24) reduces to the following

$$\frac{1}{N_t}\sum_{k=1}^{K}\frac{\gamma_k}{1+\gamma_k} \leq 1. \tag{3.25}$$

When all the UTs are demanding the same SINR $\gamma_k = \gamma\ \forall k$, the set of feasible SINR is given by $\gamma \leq \left(\frac{K}{\min\{N_t,K\}} - 1\right)^{-1}$. The result can be interpreted as follows. In the case of a single cell, as long as $N_t \geq K$, any finite SINR target is supportable. In other words, when $N_t \geq K$, the BS has enough degrees of freedom to completely eliminate the interference (possibly by zero-forcing beamforming) and transmitting with high enough power to achieve any target SINR. This condition was derived in [61] for the CBF algorithm (finite dimensional regime). In the above result, we have extended it to the asymptotic setting as well.

Corollary 2: 2-Cell Wyner Model

Consider a perfectly symmetric multi cell case where the path loss from BS_i to the UTs in cell i is 1 and the path loss to the UTs in the neighboring cell is ϵ. Every UT demands the same SINR target given by γ. In this case, (3.24) reduces to

$$\frac{K}{N_t}\left(\frac{\gamma}{1+\gamma} + \frac{\epsilon\gamma}{1+\epsilon\gamma}\right) \leq 1. \tag{3.26}$$

We would like to mention that the feasibility conditions for the special case of two cell Wyner model is derived in [60] as well.

Lastly, we show that the downlink SINR with the power allocation of the ROBF algorithm converges almost surely to the target SINR.

Theorem 6. *(Convergence of the Downlink SINR)*

$$\frac{|\mathbf{v}_{i,j}^H\mathbf{h}_{i,i,j}|^2}{\sum_{k\neq j}|\mathbf{v}_{i,k}^H\mathbf{h}_{i,i,j}|^2 + \sum_{n\neq i,k}|\mathbf{v}_{n,k}^H\mathbf{h}_{n,i,j}|^2 + N_0} \xrightarrow[N_t,K\to\infty]{a.s.} \gamma_{i,j} \qquad \forall i,j \tag{3.27}$$

The proof is provided in Section 3.7.5.

3.4.5 Numerical Results for the ROBF Algorithm

In this section, we present some numerical results to demonstrate the performance of the ROBF algorithm.

We first plot the sum downlink power for the ROBF algorithm as a function of the target SINR values for different number of transmit antennas in Figure

3.3. The cellular model considered here is a two cell Wyner model with intra-cell path loss as 1 and the inter-cell path loss 0.5. We assume that the target SINR for all the UTs is the same value. It can be seen that beyond a certain cut off value of the target SINR, the sum downlink power grows unbounded. This is precisely the value of the target SINR at which the optimization problem becomes infeasible. It can be verified that the cut off point is the one for which target SINR does not satisfy the condition in (3.26). Also note that the higher the number of transmit antennas, the higher is the cut off value of the target SINR. This is due to the availability of higher number of spatial degrees of freedom. In this case when $N_t = 2K$, any target SINR is achievable (note that $2K$ is the total number of UTs in two cells). An intuitive way to understand this is that the BSs can perform zero forcing beamforming with appropriate power allocation to achieve any target SINR.

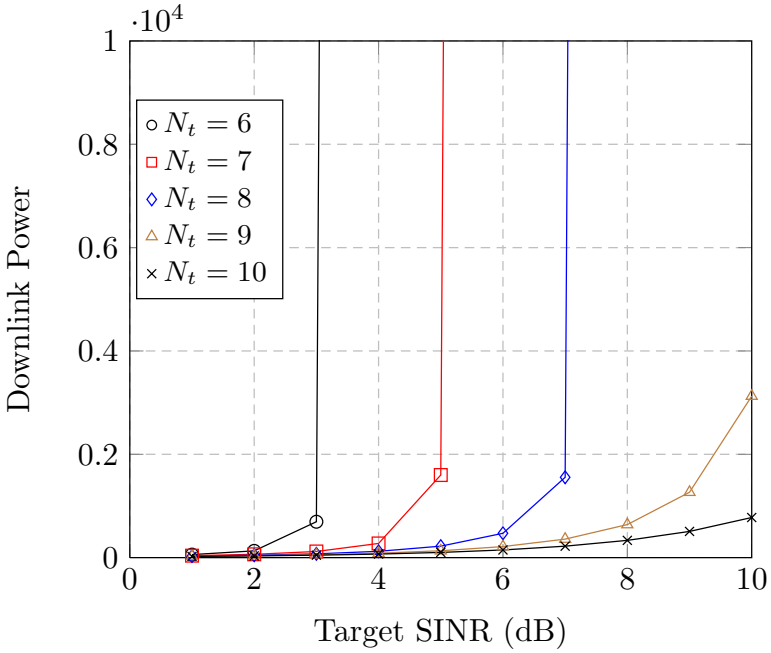

Figure 3.3: Downlink power vs target SINR for $K = 5$ UTs per cell.

Next we consider a hexagonal cellular system with a cluster of 3 cells as shown in Figure 3.4 and assume $N_t = K$. We consider a more general distance dependent path loss model in which the UTs are assumed to be arbitrarily scattered inside each cell. In this case, the path loss factor from UT k in cell i to BS j is given as $\sigma_{i,j,k} = {d_{i,j,k}}^{-\beta}$ where $d_{i,j,k}$ is the distance between UT k in cell i to BS j, normalized to the maximum distance within a cell,

and β is the path loss exponent which lies usually in the range from 2 to 5 dependent on the radio environment. We normalize the variance of the noise to $N_0 = 1$. We also assume that no user terminal is within a normalized distance of 0.1 from the closest BS. We implement the ROBF algorithm and run our simulations for 1000 channel realizations. We plot the variation of the uplink and downlink SINR (averaged across the UTs) for these 1000 channel realizations as a function of the number of UTs per cell in Figures 3.5 and 3.6, respectively. The horizontal line represents the target SINR which is 5 dB. The vertical lines represent the variation of the achieved uplink and downlink SINR around the target SINR value. It can be seen that for finite values of system dimensions, the downlink SINR fluctuates around the target SINR. As the system dimensions grow infinitely large, the SINR constraints are satisfied for the UTs for almost every given channel realization, making our algorithm asymptotically optimal.

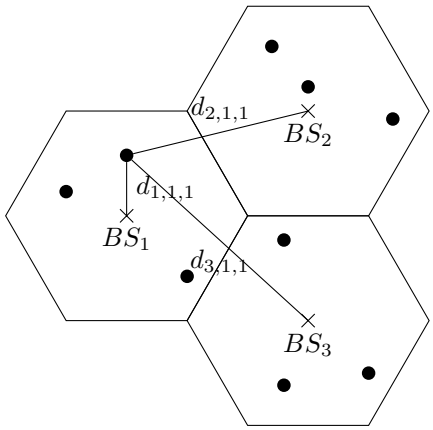

Figure 3.4: Example of a Network with 3 Cells. The crosses represent the location of the BSs and the dots represent the location of the UTs randomly scattered inside the cells. The distances of a UT from the three BSs are also provided.

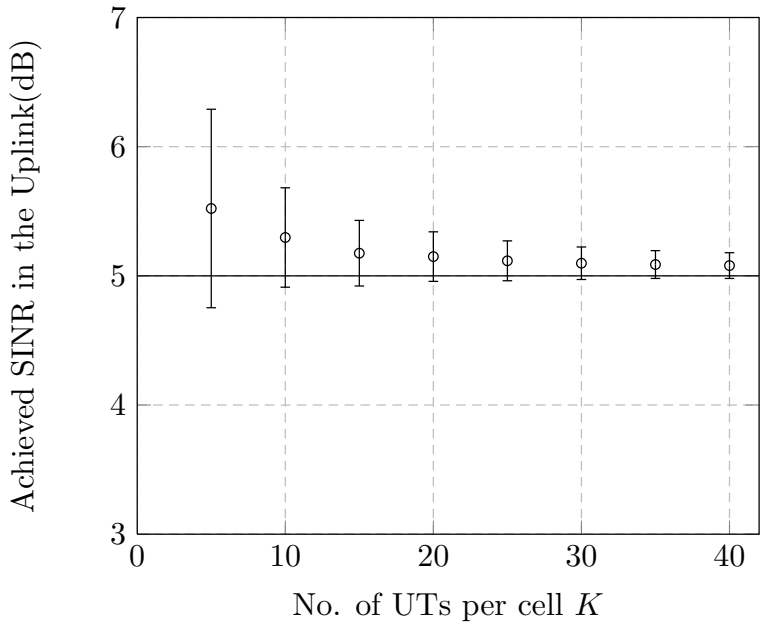

Figure 3.5: Variation of Uplink SINR as a function of the number of UTs $K, (N_t = K)$, for 1000 channel realizations. The dots represent the mean achieved uplink SINR and the vertical lines represent the standard deviation.

We now move to Part II of this chapter, where we provide the motivation and the details of the design of the flow controller.

3.5 Part II: Flow Controller Design and Queue Stability

First, we provide the motivation and the description of the flow controller design.

3.5.1 Motivation for the Flow Controller

Recall from the previous section that the actual achieved SINR in the downlink fluctuates around the target SINR. The achieved SINR in the downlink in turn depends on the channel conditions between the BSs and the UTs during the corresponding time slot. Now, the flow controller does not have the knowledge of the wireless channel conditions between the BSs and the UTs. This implies that the flow controller does not have the knowledge of the actual achieved SINR in the downlink (and hence the number of packets transmitted out of the queue

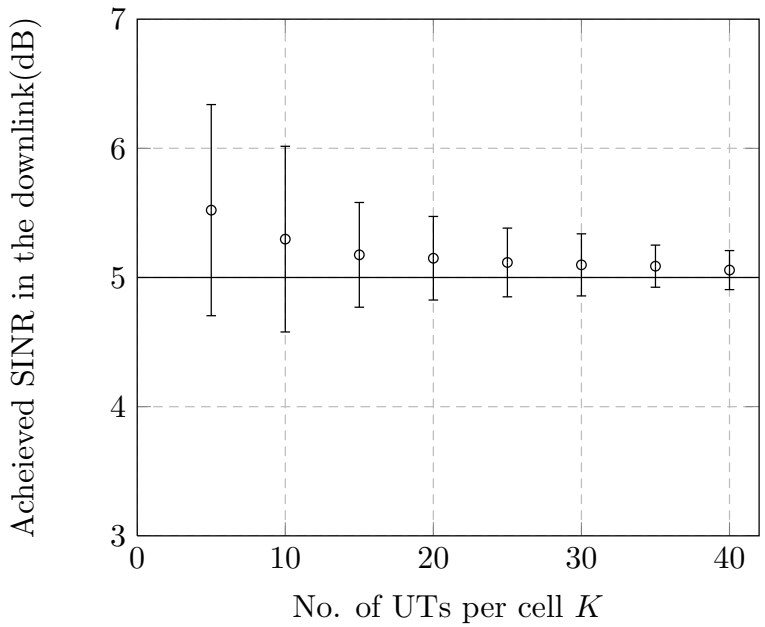

Figure 3.6: Variation of Downlink SINR as a function of the number of UTs $K, (N_t = K)$, for 1000 channel realizations. The dots represent the mean achieved downlink SINR and the vertical lines represent the standard deviation.

during the time slot t, $\mu_{i,j}^t$). Therefore, we need an effective queue-length control algorithm which can operate without the knowledge of the wireless channel conditions and minimize the fluctuations of the queue-length around the target. We model our flow controller design using the H^∞ control algorithm which we describe in the next part.

3.5.2 Flow Controller Design

We now describe the flow controller design. Recall that the equation for the queue-length evolution at the BS is given by

$$Q_{i,j}[t+1] = Q_{i,j}[t] + A_{i,j}[t] - \mu_{i,j}[t] + U_{i,j}[t] \qquad \forall i,j. \tag{3.28}$$

The task of the flow controller is to specify $A_{i,j}[t]$, the number of packets that should flow into the queue at the BS at each time slot t. $\mu_{i,j}[t]$ which is the number of packets departing the queue depends on the achieved SINR in the downlink corresponding to $\mathrm{UT}_{i,j}$. $U_{i,j}[t]$ is a constant which is added in order to keep the queue-length positive at all instants of time. For notational simplicity, we assume that the target SINR of the UTs is the same for all the time slots, i.e. $\gamma_{i,j}[t] = \bar{\gamma}_{i,j}$. Also, we denote $\bar{\mu}_{i,j} = B\log(1+\bar{\gamma}_{i,j})$. Let us subtract the target queue-length $\bar{Q}_{i,j}$ from the LHS and the RHS of (3.28). We also add and subtract the target service rate of each queue, $\bar{\mu}_{i,j}$ on the RHS of the equation (3.28). Hence equation (3.28) after rearranging becomes

$$Q_{i,j}[t] - \bar{Q}_{i,j} = [Q_{i,j}[t] - \bar{Q}_{i,j}] + [A_{i,j}[t] - \bar{\mu}_{i,j}] + [\bar{\mu}_{i,j} - \mu_{i,j}[t]] + U_{i,j}[t] \tag{3.29}$$

We perform a change of variables and define $\psi_{i,j}[t] = Q_{i,j}[t] - \bar{Q}_{i,j}$ and the vectorized version by the notation $\boldsymbol{\psi}[t] = [\psi_{1,1}[t],\ \psi_{1,2}[t], \ldots,\ \psi_{1,K}[t], \ldots,\ \psi_{N,K}[t]]^T$. We also define $u_{i,j}[t] = A_{i,j}[t] - \bar{\mu}_{i,j}$ and $\zeta[t] = (\bar{\mu}_{i,j} - \mu_{i,j}[t]) + U_{i,j}[t]$. Their vectorized notations are given by

$$\mathbf{u}[t] = [u_{1,1}[t],\ u_{1,2}[t], \ldots,\ u_{1,K}[t], \ldots,\ u_{N,K}[t]]^T$$
$$\boldsymbol{\zeta}[t] = [\zeta_{1,1}[t],\ \zeta_{1,2}[t], \ldots,\ \zeta_{1,K}[t], \ldots,\ \zeta_{N,K}[t]]^T$$

Therefore, equation (3.29) in its vectorized form becomes

$$\boldsymbol{\psi}[t+1] = \boldsymbol{\psi}[t] + \mathbf{u}[t] + \boldsymbol{\zeta}[t] \tag{3.30}$$

We assume that the flow controller has perfect information of the queue-length during every time slot t. However, the flow controller is oblivious to the achieved SINR in the downlink, and hence, the knowledge of the noise $\boldsymbol{\zeta}[t]$. Additionally,

the flow controller does not have the knowledge of the statistical distribution of the noise process.

The objective of the flow controller design problem is to minimize the cost function

$$J = \frac{1}{T}\big[\sum_{t=1}^{T}\big(||\boldsymbol{\psi}[t]||^2 + ||\mathbf{u}[t]||^2\big)\big] \tag{3.31}$$

The first term of the cost function in equation (3.31) denotes the penalty for deviating from the target queue-length. The second term tries to maintain the arrival rate at each queue as close as possible to the target service rate $\bar{\gamma}_{i,j}$.

Equation (3.30) represents a linear dynamic system with a state variable $\boldsymbol{\psi}[t]$, a control parameter $\mathbf{u}[t]$ and an unknown process noise $\boldsymbol{\zeta}[t]$. Equation (3.31) the quadratic cost to be minimized. In order to solve this problem, we use a controller based on H^∞ control algorithm with closed loop perfect state information pattern and finite observation window ([35], Theorem 3.2). The task of the controller is to specify $A_{i,j}[t]$ by calculating the control parameter $u_{i,j}[t]$ at every slot t.

The reason why H^∞ controller takes control decisions without making any assumptions on the noise process is because it tries to minimize the cost function assuming the worst case noise. For this reason, the H^∞ controller is a mini-max controller. For details on the H^∞ control design, please refer to the Section 2.3.2.

We can also model the imperfect observation of queue-length at the flow controller by adding an observation noise process. In this case the flow controller makes the observation

$$\mathbf{y}[t] = \boldsymbol{\psi}[t] + \boldsymbol{\eta}[t] \tag{3.32}$$

where $\boldsymbol{\eta}[t]$ is an observation noise parameter. Additionally, we can can also model delayed queue-length observations at the flow controller using H^∞ control with closed loop delayed information pattern ([35], Theorem 3.4). However, in order to keep the presentation simple, we do not consider the observation noise process and assume that the CS has a perfect observation of the queue-lengths.

3.5.3 Numerical Results for the Flow Controller

We now provide some numerical results for the flow controller coupled with the ROBF based beamforming algorithm. We consider the same hexagonal cellular model as in Figure 3.4 under a path loss based model. In order to show the effectiveness of the H^∞ based flow controller, we compare its performance in

minimizing the fluctuations of the queue-length around the target as opposed to two other cases. First, the case with no flow control. In this case, we assume that the number of bits flowing into the queue, $A_{i,j}[t] = \bar{\mu}_{i,j}$ (the target service rate) for all time slots t. In the second case, we use the LQG based flow control (the control decision is based on LQG) [3]. It must be recalled that the LQG control is optimal minimizing the expected cost when the noise process is Gaussian in nature. Note that if the fluctuations of the downlink SINR around the target SINR were to be indeed Gaussian, the cost incurred by the LQG controller would be lesser than that of H^∞ controller.

We assume that the target queue-length of the buffers at the BSs to be 20kBs (kilobits). We also assume the number of channel uses per coherence interval of the channel (B in equation (3.3)) as one-third the target queue-length. The channel realizations are assumed to be i.i.d. across the coherence intervals. We perform our simulations considering a target SINR of 9 dB for all the UTs during and all the time slots. We run our simulations for $T = 500$ time slots. In order to quantify the variation in the queue-lengths around the target queue size, we define the normalized mean square error (NMSE) of the queue-lengths given by

$$\mathrm{NMSE} = \frac{1}{NKT} \sum_{t=1}^{T} \sum_{i,j} \frac{(Q_{i,j}[t] - \bar{Q}_{i,j})^2}{\bar{Q}_{i,j}}.$$

We plot of NMSE of the queue-length for all the three cases in Figure 3.7. The simulation results show that the H^∞ based flow controller has a lower NMSE as compared to the case with no flow control and the LQG based flow control. This shows that when operating without the knowledge of channel conditions (and hence the number of packets transmitted out of the queue $\mu_{i,j}[t]$), a robust controller like H^∞ control performs better than the case with no control and the LQG based flow-control. It justifies our modeling of the queue-length evolution as a linear dynamical process in which the statistical distribution of the noise process is unknown.

3.6 Conclusion

In this chapter, we addressed the joint problem of traffic flow control and beamforming design in the context of SCNs. We first derived a distributed beamforming algorithm using only locally available CSI at the BSs and some statistical side information of the channel gains to other UTs. Compared to the existing works, our beamforming algorithm incurs a lower burden in terms of the information exchange between BSs and is asymptotically optimal for a large number

[3] We provide a brief description of a LQG control in section 2.3.1.

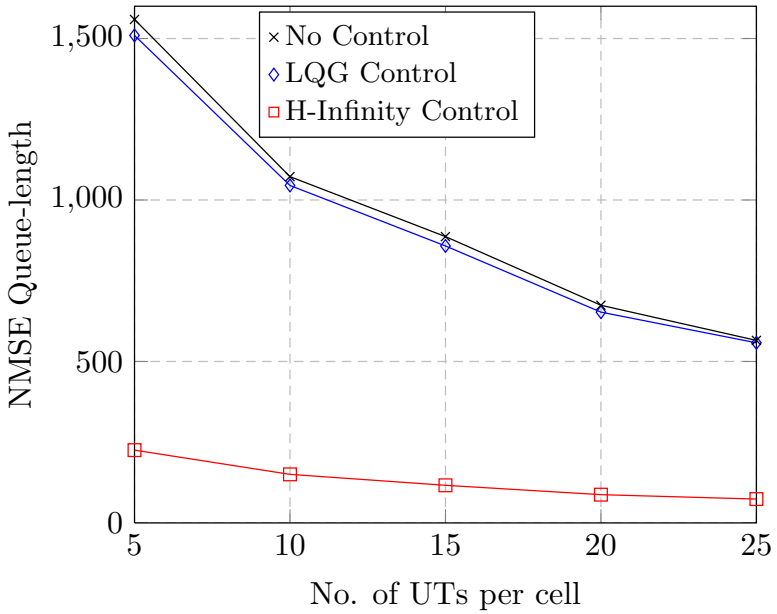

Figure 3.7: Comparison of H$^\infty$ control Vs LQG control, Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T = 500.

of BS antennas and UTs. Further on, we designed an effective H$^\infty$ based flow controller is to regulate the arrival process to the queues at the BS to minimize the fluctuations of the queue-length around the target queue-length. The H$^\infty$ based flow controller is that it functions without the knowledge of the CSI of the wireless channel between the BS and the UTs. The H$^\infty$ controller keeps the fluctuations around the target queue-length to a minimum as compared to LQG based control. A natural extension of this work would be to jointly address the two problems of flow control and beamforming design and develop a cross-layer model to address the same.

3.7 Appendices

3.7.1 Proof of the Fixed Point Equation

In order to prove the convergence of the iterative equation (3.14), we use the arguments of standard function [62]. A $K-$variate function $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \ldots, g_K(\mathbf{x})] \in \mathbb{R}^K$ for $\mathbf{x} \in \mathbb{C}^K$ is said to be standard if it fulfills the following conditions:

- Positivity: $\mathbf{g}(\mathbf{x}) > 0$, for $\mathbf{x} \geq 0$.
- Scalability: For $\beta > 1$, $\beta\mathbf{g}(\mathbf{x}) > \mathbf{g}(\beta\mathbf{x})$.

- Monotonicity: For $\mathbf{x}' \geq \mathbf{x}$, $\mathbf{g}(\mathbf{x}') \geq \mathbf{g}(\mathbf{x})$.

Now consider the following iterative algorithm given by,

$$\mathbf{x}^{p+1} = \mathbf{g}(\mathbf{x}^p), \qquad p \geq 1 \tag{3.33}$$

If the $K-$variate function is standard, then it ensures the convergence of (3.33) to its unique fixed point solution $\mathbf{x} = \mathbf{g}(\mathbf{x})$, if the solution exists.

Consider the iterative equations in (3.14) which can be represented as

$$\mu_{i,j}^{p+1} = f_{i,j}(\boldsymbol{\mu}^p) \tag{3.34}$$

where $\boldsymbol{\mu}_i^p = [\mu_{i,1}^p, \ldots, \mu_{i,K}^p]^T$ and $\boldsymbol{\mu}^p = [\boldsymbol{\mu}_i^p, \ldots, \boldsymbol{\mu}_n^p]^T$ and

$$f_{i,j}(\boldsymbol{\mu}^p) \triangleq \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^p} \qquad \forall i, j. \tag{3.35}$$

Let us also define the $NK-$ variate function

$$\begin{aligned} \mathbf{f}_i(\boldsymbol{\mu}^p) &\triangleq [f_{i,1}(\boldsymbol{\mu}^p), \ldots, f_{i,K}(\boldsymbol{\mu}^p)]^T \in \mathbb{R}^K \\ \mathbf{f}(\boldsymbol{\mu}^p) &\triangleq [\mathbf{f}_1(\boldsymbol{\mu}^p), \ldots, \mathbf{f}_N(\boldsymbol{\mu}^p)]^T \in \mathbb{R}^{NK}. \end{aligned} \tag{3.36}$$

The existence of the fixed point to the equation $\boldsymbol{\mu} = \mathbf{f}(\boldsymbol{\mu})$ follows from Lemma 1. We will now prove that the $NK-$ variate function $\mathbf{f}(\boldsymbol{\mu})$ is a standard function and hence the iterations of (3.34) converges to its unique fixed point. In the subsequent part of this proof, we will introduce the notation $\bar{m}_i^t(z)$ to denote the solution of the fixed point equation

$$\bar{m}_i^t(z) = \left(\frac{1}{N_t} \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{\sigma_{i,n,k}\mu_{n,k}^p}{1 + \sigma_{i,n,k}\mu_{n,k}^p \bar{m}_i^p(z)} + z \right)^{-1}. \tag{3.37}$$

With this notation, the solution to the fixed point equation of (3.15) can be written as $\bar{m}_i^p(1)$. For notational convenience, we will also drop the superscript p.

Positivity: The positivity result follows directly since $\bar{m}_i(1)$ is positive whenever $\boldsymbol{\mu} \geq 0$. Hence $f_{i,j}(\boldsymbol{\mu}) \geq 0$.

Scalability: Let us consider the difference between the following quantities.

$$\begin{aligned} \beta f_{i,j}(\boldsymbol{\mu}) - f_{i,j}(\beta\boldsymbol{\mu}) &= \frac{\beta\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{(1)}(1)} - \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{(2)}(1)} \\ &= \frac{\gamma_{i,j}}{\sigma_{i,i,j}} \left(\frac{\beta}{\bar{m}_i^{(1)}(1)} - \frac{1}{\bar{m}_i^{(2)}(1)} \right) \end{aligned} \tag{3.38}$$

where $\beta > 1$ and $\bar{m}_i^{(1)}(1)$ and $\bar{m}_i^{(2)}(1)$ are the unique solutions to (3.15) evaluated at $\boldsymbol{\mu}$ and $\beta\boldsymbol{\mu}$. In order to evaluate $\bar{m}_i^{(2)}(1)$, we go back to the definition $\bar{m}_i^{(2)}(1)$

evaluated at $\beta\boldsymbol{\mu}$.

$$\bar{m}_i^{(2)}(1) = \frac{1}{N}\text{tr}(\mathbf{H}_i\beta\mathbf{M}\mathbf{H}_i^H + \mathbf{I})^{-1} \asymp \frac{1}{\beta}\bar{m}_i\left(\frac{1}{\beta}\right) \tag{3.39}$$

where $\mathbf{M} = \text{diag}[\boldsymbol{\mu}]$. Clearly $1/\beta < 1$. $\bar{m}_i^{(2)}(1)$ can be evaluated as the solution to the fixed point equation (3.15) evaluated at the point $1/\beta$ and then scaling the result by β. From (3.38) and (3.39), it can be concluded that in order to prove the scalability result, it is sufficient to show that $\bar{m}_i(z)$ is a decreasing function of z.

In order to prove the same, let us consider an extended version of the channel matrix which is constructed as follows. Defining, $\mathbf{R}_{i,j,k} = \sigma_{i,j,k}\mathbf{I}_{N_t} \in \mathbb{R}^{N_t\times N_t}$ and $\mathbf{R}^L_{i,j,k} = \sigma_{i,j,k}\mathbf{I}_{N_tL} \in \mathbb{R}^{N_tL\times N_tL}$. The matrix $\mathbf{H}_i^L$ is constructed as follows:

$$\begin{aligned}\mathbf{H}_{i,j}^L &= \frac{1}{\sqrt{L}}\left[\mathbf{R}_{i,j,1}^{1/2}\mathbf{X}_{i,j}^L, \ldots, \mathbf{R}_{i,j,K}^{1/2}\mathbf{X}_{i,j}^L\right] \in \mathbb{C}^{N_tL\times KL}\\ \mathbf{H}_i^L &= \left[\mathbf{H}_{i,1}^L, \ldots, \mathbf{H}_{i,N}^L\right] \in \mathbb{C}^{N_tL\times NKL}\end{aligned} \tag{3.40}$$

where the matrix $\mathbf{X}_{i,j}^L \in \mathbb{C}^{Nt\times K}$, whose elements $\mathbf{X}_{i,j}^L(p,q) \sim \mathcal{CN}(0, \frac{1}{N_t})$. Also, let us define the following,

$$\begin{aligned}\boldsymbol{\lambda}_{i,j}^L &= \left[\lambda_{i,j}, \ldots, \lambda_{i,j}\right]^T \in \mathbb{R}^L\\ \boldsymbol{\lambda}_i^L &= \left[\boldsymbol{\lambda}_{i,1}^L, \ldots, \boldsymbol{\lambda}_{i,K}^L\right]^T \in \mathbb{R}^{KL}\\ \boldsymbol{\Lambda}^L &= \text{diag}\left(\boldsymbol{\lambda}_1^L, \ldots, \boldsymbol{\lambda}_N^L\right) \in \mathbb{R}^{NKL\times NKL}\end{aligned} \tag{3.41}$$

and let $\mathbf{Q}_i(z) = (\mathbf{H}_i\boldsymbol{\Lambda}\mathbf{H}_i^H + z\mathbf{I}_{N_tL})^{-1}$ and $\mathbf{Q}_i^L(z) = (\mathbf{H}_i^L\boldsymbol{\Lambda}^L(\mathbf{H}^L)_i^H + z\mathbf{I}_{N_tL})^{-1}$. Let us denote

$$m_i^L(z) = \frac{1}{N_tL}\text{tr}(\mathbf{Q}_i^L(z)). \tag{3.42}$$

It can be verified that for any fixed N, N_t and K, the following limit holds,

$$m_i^L(z) - \bar{m}_i(z) \xrightarrow[L\to\infty]{\text{a.s.}} 0. \tag{3.43}$$

Now consider the difference between the following two quantities, for any $z_2 > z_1 > 0$ and for any positive L, we have

$$\begin{aligned}&m_i^L(z_1) - m_i^L(z_2)\\ &= \frac{1}{N_tL}\text{tr}\left(\mathbf{Q}_i^L(z_1) - \mathbf{Q}_i^L(z_2)\right)\\ &\overset{(a)}{>} 0\end{aligned} \tag{3.44}$$

where inequality (a) follows by the following identity: For invertible matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{B}^{-1}. \tag{3.45}$$

We will now show that for any fixed N, N_t and K, the inequality in (3.44) of the random quantities $(m_i^L(z))$ also hold for their respective deterministic approximations $(\bar{m}_i(z))$. This can be argued as follows. Consider the difference

$$\begin{aligned} m_i^L(z_1) - m_i^L(z_2) = m_i^L(z_1) - \bar{m}_i(z_1) - m_i^L(z_2) \\ + \bar{m}_i(z_2) + \bar{m}_i(z_1) - \bar{m}_i(z_2). \end{aligned} \tag{3.46}$$

First note that since the matrix $\mathbf{H}_i^L$ has bounded spectral norm almost surely, it follows that, almost surely,

$$\lim_{L\to\infty} m_i^L(z) > 0. \tag{3.47}$$

From (3.47), further applying the result of (3.43) in the right hand side of (3.46) and the inequality of (3.44), it follows that

$$\bar{m}_i(z_1) - \bar{m}_i(z_2) > 0. \tag{3.48}$$

Thus, $\bar{m}_i(z)$ is a decreasing function of z. We remark that the strict positivity of the $m_i^L(z)$ in (3.47) is essential for the strict positivity argument of (3.48).

Monotonicity: Consider $\boldsymbol{\mu}' \geq \boldsymbol{\mu}$. In this case, we denote $\bar{m}_i^{(1)}(1)$ and $\bar{m}_i^{(2)}(1)$ as the solutions to the fixed point equations in (3.15) evaluated at $\boldsymbol{\mu}'$ and $\boldsymbol{\mu}$ respectively. As before, let us consider the difference between the quantities,

$$f_{i,j}(\boldsymbol{\mu}') - f_{i,j}(\boldsymbol{\mu}) = \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{(1)}(1)} - \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{(2)}(1)}. \tag{3.49}$$

We now have to show that $\bar{m}_i^{(1)}(1) \leq \bar{m}_i^{(2)}(1)$ in order to prove the monotonicity result. This can be shown by constructing the extended matrices as in (3.40) and noting the monotonicity property for the associated random quantities and extending the result to their deterministic approximations for any system dimensions. The proof is similar to the scalability result and hence omitted here.

3.7.2 Computation of $\bar{\boldsymbol{\delta}}$

Let us recall the downlink SINR constraint for the optimization problem with the solution provided by the ROBF algorithm

$$\frac{\bar{\delta}_{i,j}|\hat{\mathbf{v}}_{i,j}^H \mathbf{h}_{i,i,j}|^2}{\sum_{k\neq j}\bar{\delta}_{i,k}|\hat{\mathbf{v}}_{i,k}^H \mathbf{h}_{i,i,j}|^2 + \sum_{n\neq i,k}\bar{\delta}_{n,k}|\hat{\mathbf{v}}_{n,k}^H \mathbf{h}_{n,i,j}|^2 + N_0} \geq \gamma_{i,j}. \tag{3.50}$$

At the optimal point, all the SINR constraints must be satisfied with equality [42]. Hence,

$$\frac{\bar{\delta}_{i,j}}{\gamma_{i,j}}|\hat{\mathbf{v}}_{i,j}^H\mathbf{h}_{i,i,j}|^2 - \sum_{k\neq j}\bar{\delta}_{i,k}|\hat{\mathbf{v}}_{i,k}^H\mathbf{h}_{i,i,j}|^2 - \sum_{n\neq i,k}\bar{\delta}_{n,k}|\hat{\mathbf{v}}_{n,k}^H\mathbf{h}_{n,i,j}|^2 = N_0 \ \forall i,j \tag{3.51}$$

The last equation provides us with a set of NK linear equations in $\bar{\boldsymbol{\delta}}$ which can be solved using matrix inversion, in case a unique solution exists.

We will now seek to find the deterministic equivalent of the coefficients of the linear equations. Let us first focus on the expression for the useful signal power. Using Lemma 2, Lemma 7 and Theorem 3 it can be shown that

$$|\hat{\mathbf{v}}_{i,j}^H\mathbf{h}_{i,i,j}|^2 \asymp \left(\frac{\sigma_{i,i,j}\bar{m}_i}{1+\mu_{i,j}\sigma_{i,i,j}\bar{m}_i}\right)^2. \tag{3.52}$$

Note that these steps are standard and can be found e.g. in the textbook [58]. We now find the deterministic equivalent of the interference power. First let us focus on the intra-cell interference term.

$$\begin{aligned}
|\hat{\mathbf{v}}_{i,k}^H\mathbf{h}_{i,i,j}|^2 &= \mathbf{h}_{i,i,j}^H\left(\boldsymbol{\Sigma}_i^{\mu}+\mathbf{I}\right)^{-1}\mathbf{h}_{i,i,k}\mathbf{h}_{i,i,k}^H\left(\boldsymbol{\Sigma}_i^{\mu}+\mathbf{I}\right)^{-1}\mathbf{h}_{i,i,j} \\
&\asymp \frac{\bar{G}_{i,i,j}}{N_t}\mathrm{tr}\left(\boldsymbol{\Sigma}_i'^{\mu}+\mathbf{I}\right)^{-1}\mathbf{h}_{i,i,k}\mathbf{h}_{i,i,k}^H\left(\boldsymbol{\Sigma}_i'^{\mu}+\mathbf{I}\right)^{-1} \\
&= \frac{\bar{G}_{i,i,j}}{N_t}\mathrm{tr}\left(\mathbf{h}_{i,i,k}^H\left(\boldsymbol{\Sigma}_i'^{\mu}+\mathbf{I}\right)^{-2}\mathbf{h}_{i,i,k}\right) \\
&\asymp \frac{1}{N_t}\bar{G}_{i,i,j}\bar{G}_{i,i,k}\bar{m}_i'.
\end{aligned} \tag{3.53}$$

Similar to the arguments made in (3.53), the deterministic equivalent for the inter-cell interference term can be given by

$$|\hat{\mathbf{v}}_{n,k}^H\mathbf{h}_{n,i,j}|^2 \asymp \frac{1}{N_t}\bar{G}_{n,i,j}\bar{G}_{n,n,k}\bar{m}_n'. \tag{3.54}$$

We now replace each term of (3.51) by its asymptotic approximation to obtain

$$\frac{1}{\gamma_{i,j}}\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2\bar{\delta}_{i,j} - \sum_{\substack{k=1\\k\neq j}}^{K}\frac{1}{N_t}\bar{G}_{i,i,j}\bar{G}_{i,i,k}\bar{m}_i'\bar{\delta}_{i,k} - \sum_{n\neq i,k}\frac{1}{N_t}\bar{G}_{n,i,j}\bar{G}_{n,n,k}\bar{m}_n'\bar{\delta}_{n,k} = N_0. \tag{3.55}$$

Rearranging the terms to solve for $\bar{\delta}_{i,j}$ leads to

$$\bar{\delta}_{i,j} - \frac{\sum_{\substack{k=1\\k\neq j}}^{K}\gamma_{i,j}\bar{G}_{i,i,j}\bar{G}_{i,i,k}\bar{m}_i'\bar{\delta}_{i,k}}{N_t\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2} - \frac{\sum_{n\neq i,k}\gamma_{i,j}\bar{G}_{n,i,j}\bar{G}_{n,n,k}\bar{m}_n'\bar{\delta}_{n,k}}{N_t\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2} = \frac{N_0}{\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2}. \tag{3.56}$$

Equation (3.56) can be rewritten in matrix form as

$$(\mathbf{I} - \boldsymbol{\Gamma}\boldsymbol{\Delta})\bar{\boldsymbol{\delta}} = \boldsymbol{\rho} \tag{3.57}$$

where the notations $\boldsymbol{\Delta}, \boldsymbol{\delta}$ and $\boldsymbol{\rho}$ are defined in the description of Algorithm 2. Hence, if the matrix $\mathbf{I}-\boldsymbol{\Gamma}\boldsymbol{\Delta}$ is invertible (which will be shown later in subsection 3.7.4),

$$\bar{\boldsymbol{\delta}} = (\mathbf{I} - \boldsymbol{\Gamma}\boldsymbol{\Delta})^{-1}\boldsymbol{\rho}. \tag{3.58}$$

3.7.3 Proof of Lemma 1

Recall that the expression of uplink SINR with power allocation $\mu_{i,j}$ is given in (3.22). $\mu_{i,j}$ is independent of the elements of the channel matrix. Also note that since $\mu_{i,j}$ is the solution to the fixed point equation (3.14), it must be bounded from above. Applying Lemma 6 to the quadratic term of $\mathbf{h}_{i,i,j}^H(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j}$ yields to,

$$\mathbb{E}\left[\left|\mathbf{h}_{i,i,j}^H(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j} - \frac{1}{N_t}\mathrm{tr}(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\right|^k\right] \leq \frac{C_1}{N_t^{\frac{k}{2}}} \qquad \forall i,j \tag{3.59}$$

for $k \geq 2$, and constant C_1 independent of N_t and K. Additionally, from the result of Lemma 6.1, [63], we have

$$\mathbb{E}\left[\left|\frac{1}{N_t}\mathrm{tr}(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1} - \bar{m}_i\right|^k\right] \leq \frac{C_2}{N_t^{\frac{k}{2}}},\ \forall i,j \tag{3.60}$$

for $k \geq 2$ constant C_2 independent of N_t and K. Therefore, from (3.59), (3.60) and Holder's inequality ($|x+y|^k \leq 2^{k-1}(|x|^k + |y|^k)$) we conclude that for some constant C_3,

$$\mathbb{E}\left[\left|\mathbf{h}_{i,i,j}^H(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j} - \bar{m}_i\right|^k\right] \leq \frac{C_3}{N_t^{\frac{k}{2}}},\ \forall i,j. \tag{3.61}$$

From (3.22), we have, $\mathbf{h}_{i,i,j}^H(\boldsymbol{\Sigma}_i'^{\mu} + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j} = \frac{\Lambda_{i,j}(\mu_{i,j})}{\sigma_{i,i,j}\mu_{i,j}}$. Moreover, at the convergence of the fixed point equation (3.14), $\bar{m}_i = \frac{\gamma_{i,j}}{\sigma_{i,i,j}\mu_{i,j}}$. Substituting in (3.61) we have,

$$\mathbb{E}\left[\left|\Lambda_{i,j}(\mu_{i,j}) - \gamma_{i,j}\right|^k\right] \leq \frac{C_4}{N_t^{\frac{k}{2}}},\ \forall i,j \tag{3.62}$$

where $C_4 = C_3\sigma_{i,i,j}\mu_{i,j}$. Note that $\sigma_{i,i,j}$ is bounded. For now, we will assume that $\mu_{i,j}$ is bounded (and in specific $\limsup_{Nt,K\to\infty}\mu_{i,j}$ is bounded as well). In

fact, later in Proposition 1, we will characterize the feasible set of target SINRs $(\gamma_{i,j})$ and show that for any feasible target SINR, $\limsup_{N_t,K\to\infty} \mu_{i,j}$ is bounded as well. In order to prove convergence results, we examine the properties of the supremum over all the indices i, j.

$$\mathbb{E}\left[\sup_{i,j}\left|\Lambda_{i,j}(\mu_{i,j})-\gamma_{i,j}\right|^k\right] \overset{(a)}{\leq} \sum_{i,j}\mathbb{E}\left[\left|\Lambda_{i,j}(\mu_{i,j})-\gamma_{i,j}\right|^k\right] \overset{(b)}{\leq} NK\frac{C_4}{N_t^{\frac{k}{2}}} = \frac{C_5}{N_t^{\frac{k}{2}-1}} \tag{3.63}$$

where $C_5 = C_4 N\beta$ (where $\beta = \frac{K}{N_t}$, a finite value). Inequality (a) follows from the linearity of expectation operation and (b) follows from the bound in (3.62). Finally, we make use of the following inequality (which can be shown easily, details omitted here)

$$\sup_{i,j}\mathbb{E}\left[|\Lambda_{i,j}(\mu_{i,j})-\gamma_{i,j}|^k\right] \leq \mathbb{E}\left[\sup_{i,j}|\Lambda_{i,j}(\mu_{i,j})-\gamma_{i,j}|^k\right]. \tag{3.64}$$

From (3.63) and (3.64) we deduce,

$$\sup_{i,j}\mathbb{E}\left[|\Lambda_{i,j}(\mu_{i,j})-\gamma_{i,j}|^k\right] \leq \frac{C_5}{N_t^{\frac{k}{2}-1}}. \tag{3.65}$$

By taking k to be sufficiently high ($k \geq 6$ in this case), the right hand side of (3.65) is summable. By Markov Inequality ((5.31) of [31]) and the Borel Cantelli lemma (Theorem 4.3 of [31]), it follows that $\Lambda_{i,j}(\mu_{i,j}) - \gamma_{i,j} \xrightarrow[N_t,K\to\infty]{a.s.} 0,\ \forall i,j$.

3.7.4 Feasibility Conditions

Part I: Proof of Lemma 2

In order to prove the lemma, let us first consider the expression for the uplink SINR for the ROBF algorithm given by

$$\Lambda_{i,j}(\mu_{i,j}) = \frac{\mu_{i,j}|\hat{\mathbf{v}}_{i,j}^H\mathbf{h}_{i,i,j}|^2}{\sum_{(n,k)\neq(i,j)}\mu_{n,k}|\hat{\mathbf{v}}_{i,j}^H\mathbf{h}_{i,n,k}|^2 + ||\hat{\mathbf{v}}_{i,j}||_2^2}. \tag{3.66}$$

For the uplink MMSE beamforming vectors given by (3.9), the expression for the uplink SINR can be given in alternate form as

$$\Lambda_{i,j}(\mu_{i,j}) = \mu_{i,j}\mathbf{h}_{i,i,j}^H(\mathbf{\Sigma}_i' + \mathbf{I}_{N_t})^{-1}\mathbf{h}_{i,i,j}. \tag{3.67}$$

Therefore, for the ROBF algorithm, the right hand side of (3.66) and (3.67) are equivalent. Let us consider the same in the asymptotic domain. Considering the deterministic equivalents for the terms of the numerator and the denominator of

the right hand side of (3.66) and following similar steps as in subsection 3.7.2, we have ($\bar{G}$ terms defined in (3.19))

$$\Lambda_{i,j}(\mu_{i,j}) \asymp \frac{\mu_{i,j}\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2}{\sum_{(n,k)\neq(i,j)} \frac{1}{N_t}\mu_{n,k}\bar{G}_{i,n,k}\bar{G}_{i,i,j}\bar{m}_i' + \alpha_i\bar{G}_{i,i,j}\bar{m}_i'} \tag{3.68}$$

$$= \frac{\mu_{i,j}\sigma_{i,i,j}\bar{m}_i^2}{\sum_{(n,k)\neq(i,j)} \frac{1}{N_t}\mu_{n,k}\bar{G}_{i,n,k}\bar{m}_i' + \alpha_i\bar{m}_i'}. \tag{3.69}$$

We now consider a slightly modified version of the expression in the right hand side of (3.69) given by

$$\frac{\mu_{i,j}\sigma_{i,i,j}\bar{m}_i^2}{\sum_{n,k} \frac{1}{N_t}\mu_{n,k}\bar{G}_{i,n,k}\bar{m}_i' + \alpha_i\bar{m}_i'}. \tag{3.70}$$

Similarly, considering the deterministic equivalents for the right hand side of (3.67),

$$\Lambda_{i,j}(\mu_{i,j}) \asymp \mu_{i,j}\sigma_{i,i,j}\bar{m}_i. \tag{3.71}$$

In what follows, we show that the expressions of (3.70) and the right hand side of (3.71) are equivalent.

Recall the fixed point equation for the computation of $\bar{m}_i$ in (3.15). Upon rearranging the terms, we have

$$\bar{m}_i = 1 - \frac{1}{N_t}\sum_{n,k} \frac{\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}{1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}. \tag{3.72}$$

Substituting the expressions for $\bar{G}_{i,n,k}$, $\bar{m}_i$ and $\bar{m}_i'$ in (3.70), we get

$$(3.70) = \frac{\sigma_{i,i,j}\mu_{i,j}\left(1 - \frac{1}{N_t}\sum_{n,k}\frac{(\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}\right)}{1+\frac{1}{N_t}\sum_{n,k}\frac{\sigma_{i,n,k}\mu_{n,k}}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}}. \tag{3.73}$$

Multiplying and diving by $\bar{m}_i$ in (3.73), we obtain

$$\begin{aligned}
(3.70) &= \frac{\sigma_{i,i,j}\mu_{i,j}\bar{m}_i\left(1 - \frac{1}{N_t}\sum_{n,k}\frac{(\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}\right)}{\bar{m}_i+\frac{1}{N_t}\sum_{n,k}\frac{\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}} \\
&= \frac{\sigma_{i,i,j}\mu_{i,j}\bar{m}_i\left(1 - \frac{1}{N_t}\sum_{n,k}\frac{(\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}\right)}{1-\frac{1}{N_t}\sum_{n,k}\frac{\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}{1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}+\frac{1}{N_t}\sum_{n,k}\frac{\sigma_{i,n,k}\mu_{n,k}\bar{m}_i}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}} \\
&= \frac{\sigma_{i,i,j}\mu_{i,j}\bar{m}_i\left(1 - \frac{1}{N_t}\sum_{n,k}\frac{(\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}\right)}{1-\frac{1}{N_t}\sum_{n,k}\frac{(\sigma_{i,n,k}\mu_{n,k}m_i)^2}{(1+\sigma_{i,n,k}\mu_{n,k}\bar{m}_i)^2}} \\
&= \sigma_{i,i,j}\mu_{i,j}\bar{m}_i.
\end{aligned} \tag{3.74}$$

This above form is exactly the same form as in (3.71). Therefore, asymptotically the two forms of SINR are equivalent. At the convergence of the fixed point

equation for evaluating the uplink power allocation of the ROBF algorithm,

$$\begin{aligned}\gamma_{i,j} &= \mu_{i,j}\sigma_{i,i,j}\bar{m}_i \\ &= \frac{\mu_{i,j}\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2}{\sum_{n,k}\mu_{n,k}\bar{G}_{i,n,k}\bar{G}_{i,i,j}\bar{m}_i' + \bar{G}_{i,i,j}\bar{m}_i'}.\end{aligned}$$

Rearranging, we have

$$\begin{aligned}\frac{1}{\gamma_{i,j}}\mu_{i,j}\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2 &= \sum_{n,k}\mu_{n,k}\bar{G}_{i,n,k}\bar{G}_{i,i,j}\bar{m}_i' \\ &\quad + \bar{G}_{i,i,j}\bar{m}_i' \\ \implies \mu_{i,j} &= \sum_{n,k}\gamma_{i,j}\frac{\mu_{n,k}\bar{G}_{i,n,k}\bar{G}_{i,i,j}\bar{m}_i'}{\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2} \\ &\quad + \frac{\bar{G}_{i,i,j}\bar{m}_i'}{\sigma_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i^2}.\end{aligned}$$

The equation in matrix form can be written as

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}')^T\boldsymbol{\mu} + \kappa \tag{3.75}$$

where $\kappa_i = \left[\frac{\bar{m}_i'}{\sigma_{i,i,1}\bar{m}_i^2}, \ldots, \frac{\bar{m}_i'}{\sigma_{i,i,K}\bar{m}_i^2}\right]^T$ and $\kappa = [\kappa_1, \ldots, \kappa_N]^T$. The matrices $\boldsymbol{\Gamma}$ is defined in (3.19). The matrix $\boldsymbol{\Delta}'$ is defined as follows.

$$\boldsymbol{\Delta}' = \begin{pmatrix} (\boldsymbol{\Delta}')^{1,1} & \ldots & (\boldsymbol{\Delta}')^{1,N} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{\Delta}')^{N,1} & \ldots & (\boldsymbol{\Delta}')^{N,N} \end{pmatrix} \tag{3.76}$$

where each submatrix $(\boldsymbol{\Delta}')^{i,j} \in \mathbb{C}^{K\times K}$ is given by

$$(\boldsymbol{\Delta}')^{i,n}_{j,k} \triangleq \begin{cases} \frac{1}{N_t}\bar{G}_{i,i,j}\bar{G}_{i,i,j}\bar{m}_i', & n = i,\ k = j \\ \frac{1}{N_t}\bar{G}_{i,i,j}\bar{G}_{i,i,k}\bar{m}_i', & n = i,\ k \neq j \\ \frac{1}{N_t}\bar{G}_{n,i,j}\bar{G}_{n,n,k}\bar{m}_n', & n \neq i. \end{cases} \tag{3.77}$$

Notice that if the uplink problem is feasible, the linear equations in (3.75) must have a solution. In other words, the matrix $\mathbf{I} - \boldsymbol{\Gamma}(\boldsymbol{\Delta}')^T$ must be invertible. Invertibility of the matrix implies that $\rho(\boldsymbol{\Gamma}(\boldsymbol{\Delta}')^T) < 1$ (where ρ is the spectral radius of the matrix). Now recall the matrix $\boldsymbol{\Delta}$ defined in equation (3.19). Observe that the matrix $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}'$ only differ in the diagonal element. Also, we can note that $\boldsymbol{\Delta} \leq \boldsymbol{\Delta}'$ and hence $\boldsymbol{\Gamma}\boldsymbol{\Delta} \leq \boldsymbol{\Gamma}\boldsymbol{\Delta}'$. Hence, it follows from Lemma 4, that $\rho(\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T) \leq \rho(\boldsymbol{\Gamma}(\boldsymbol{\Delta}')^T) < 1$.

Now recall the linear equations for computing $\bar{\boldsymbol{\delta}}$ given by

$$\bar{\boldsymbol{\delta}} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\bar{\boldsymbol{\delta}} + \boldsymbol{\rho}.$$

It has been established in works before [64], the eigen values of the matrices $\mathbf{\Gamma\Delta}$ and $\mathbf{\Gamma\Delta}^T$ are the same. For completeness, this can be argued as follows,

$$\begin{aligned}|\mathbf{\Gamma\Delta} - \lambda\mathbf{I}| &= |\mathbf{\Gamma}||\mathbf{\Delta} - \lambda\mathbf{\Gamma}^{-1}| = |\mathbf{\Gamma}|| \left(\mathbf{\Delta} - \lambda\mathbf{\Gamma}^{-1}\right)^T| \\ &= |\mathbf{\Gamma}||\mathbf{\Delta}^T - \lambda\mathbf{\Gamma}^{-1}| = |\mathbf{\Gamma\Delta}^T - \lambda\mathbf{I}| = 0.\end{aligned}$$

From the above discussion, we conclude that the spectral radius of $\mathbf{\Gamma\Delta}$ and $\mathbf{\Gamma\Delta}^T$ are the same and, hence, if $\rho(\Gamma\mathbf{\Delta}) < 1$ then $\rho(\mathbf{\Gamma\Delta}^T) < 1$ and vice versa. Consequently, the matrix $\mathbf{I} - \mathbf{\Gamma\Delta}$ is invertible. That is, the feasibility of uplink and downlink are equivalent for the ROBF problem and it is sufficient to look at the feasibility of the uplink power allocation problem alone.

Part II: Feasibility conditions for the modified system

In order to obtain the feasibility conditions, let us now focus on the uplink power allocation problem. Recall that the equation relating the target SINR and the uplink power allocation $\mu_{i,j}$ given by,

$$\gamma_{i,j} = \sigma_{i,i,j}\mu_{i,j}\bar{m}_i \qquad \forall i, j. \tag{3.78}$$

Consider the power allocation $\mu_{i,j}$ which is the solution (3.78). Notice from (3.78) that for the uplink power allocation problem to be feasible, the value of $\bar{m}_i$ must be positive. For the *modified system* $\sigma_{i,n,k} = \sigma_{max}(n), n \neq i$ (recall that where $\sigma_{max}(n)$ is the path loss of the channel between BS in cell i and the edge of the cell which is at a distance n cells from cell i). Therefore, for the *modified system* (3.15) the following must hold

$$\begin{aligned}\frac{1}{\bar{m}_i} = &\frac{1}{N_t}\sum_{k=1}^{K}\frac{\sigma_{i,i,k}\mu_{i,k}}{1+\sigma_{i,i,k}\mu_{i,k}\bar{m}_i} \\ &+ \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{max}(n)\mu_{n,k}}{1+\sigma_{max}(n)\mu_{n,k}\bar{m}_i} + 1.\end{aligned}$$

Rearranging, we have

$$\begin{aligned}1 &= \frac{1}{N_t}\sum_{k=1}^{K}\frac{\sigma_{i,i,k}\mu_{i,k}\bar{m}_i}{1+\sigma_{i,i,k}\mu_{i,k}\bar{m}_i} \\ &\qquad + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{max}(n)\mu_{n,k}}{1+\sigma_{max}(n)\mu_{n,k}\bar{m}_i} + \bar{m}_i \\ &= \frac{1}{N_t}\sum_{k=1}^{K}\frac{\gamma_{i,k}}{1+\gamma_{i,k}} \\ &\qquad + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{max}(n)\mu_{n,k}\bar{m}_i}{1+\sigma_{max}(n)\mu_{n,k}\bar{m}_i} + \bar{m}_i. \end{aligned} \tag{3.79}$$

Equation (3.79) is true for all $\bar{m}_i, i = 1, \ldots, N$. Therefore, by symmetry of the fixed point equation $\bar{m}_i = \bar{m}, i = 1, \ldots, N$. Thus,

$$\sigma_{max}(n)\mu_{n,k}\bar{m}_i = \sigma_{max}(n)\mu_{n,k}\bar{m} = \frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}. \tag{3.80}$$

Substituting (3.79) in (3.80) yields

$$1 = \frac{1}{N_t}\sum_{k=1}^{K}\frac{\gamma_{i,k}}{1+\gamma_{i,k}} + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}}{1+\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}} + \bar{m}_i.$$

Considering the asymptotic regime, taking $\liminf_{Nt,K\to\infty}$, yields,

$$\begin{aligned}\liminf_{Nt,K\to\infty}\Big[\frac{1}{N_t}\sum_{k=1}^{K}\frac{\gamma_{i,k}}{1+\gamma_{i,k}} + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}}{1+\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}} \\ +\bar{m}_i\Big] = 1.\end{aligned} \tag{3.81}$$

Noting that

$$\liminf_{n\to\infty}[a_n + b_n] \leq \limsup_{n\to\infty} a_n + \liminf_{n\to\infty} b_n,$$

we have

$$\begin{aligned}\limsup_{Nt,K\to\infty}\Big[\frac{1}{N_t}\sum_{k=1}^{K}\frac{\gamma_{i,k}}{1+\gamma_{i,k}} + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}}{1+\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}}\Big] \\ + \liminf_{N_t,K\to\infty}\bar{m}_i \geq 1.\end{aligned} \tag{3.82}$$

Now, recall the relation from the relation

$$\mu_{i,j} = \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i}.$$

The feasibility of the uplink problem implies $\limsup_{Nt,K\to\infty} \mu_{i,j} < \infty$. This implies that $\liminf_{Nt,K\to\infty} \bar{m}_i > 0$ (strictly positive). Therefore, from equation (3.82), if follows that the feasibility conditions are given by,

$$\limsup_{Nt,K\to\infty} \left[\frac{1}{N_t} \sum_{k=1}^{K} \frac{\gamma_{i,k}}{1+\gamma_{i,k}} + \frac{1}{N_t} \sum_{\substack{n=1 \\ n\neq i}}^{N} \sum_{k=1}^{K} \frac{\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}}{1+\frac{\sigma_{max}(n)}{\sigma_{n,n,k}}\gamma_{n,k}} \right] < 1 \qquad \forall i. \tag{3.83}$$

Part III: Proof of Lemma 3

Assume the feasible target SINR $\gamma_{i,j},\ \forall i,j$. We now solve the downlink power minimization problem with the ROBF algorithm under two system models, namely the original system and the modified system.

First consider the solution of the original system. The fixed point equation for the computation of the uplink power allocation must satisfy the following equations (we use the superscript "org" to represent the original system)

$$\gamma_{i,j} = \sigma_{i,i,j}\mu_{i,k}^{\text{org}}\bar{m}_i^{\text{org}} \tag{3.84}$$

$$\frac{1}{\bar{m}_i^{\text{org}}} = \frac{1}{N_t}\frac{\sigma_{i,i,k}\mu_{i,k}^{\text{org}}}{1+\sigma_{i,i,k}\mu_{i,k}^{\text{org}}\bar{m}_i^{\text{org}}} + \frac{1}{N_t}\sum_{\substack{n=1 \\ n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{i,n,k}\mu_{n,k}^{\text{org}}}{1+\sigma_{i,n,k}\mu_{n,k}^{\text{org}}\bar{m}_i^{\text{org}}} + 1. \tag{3.85}$$

Similarly, the fixed point equation for the *modified system* must satisfy the following equations (we use the superscript "mod" to represent the modified system)

$$\gamma_{i,j} = \sigma_{i,i,j}\mu_{i,k}^{\text{mod}}\bar{m}_i^{\text{mod}} \tag{3.86}$$

$$\frac{1}{\bar{m}_i^{\text{mod}}} = \frac{1}{N_t}\frac{\sigma_{i,i,k}\mu_{i,k}^{\text{mod}}}{1+\sigma_{i,i,k}\mu_{i,k}^{\text{mod}}\bar{m}_i^{\text{mod}}} + \frac{1}{N_t}\sum_{\substack{n=1 \\ n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{max}(n)\mu_{n,k}^{\text{mod}}}{1+\sigma_{max}(n)\mu_{n,k}^{\text{mod}}\bar{m}_i^{\text{mod}}} + 1. \tag{3.87}$$

We will now show that any feasible SINR for the *modified system* is achievable in the original system. Let us denote $\boldsymbol{\mu}_i^{\text{mod}} = [\mu_{i,1}^{\text{mod}}, \ldots, \mu_{i,K}^{\text{mod}}]^T$ and $\boldsymbol{\mu}^{\text{mod}} = [\boldsymbol{\mu}_1^{\text{mod}}, \ldots, \boldsymbol{\mu}_N^{\text{mod}}]^T$. The main idea behind showing this is to prove that for given target SINR of $\gamma_{i,j}$, $\boldsymbol{\mu}^{\text{mod}}$ is a feasible power allocation in the original system, i.e., when we use the power allocation $\boldsymbol{\mu}^{\text{mod}}$ in the original system, the UTs achieve at least their SINR targets. Firstly, it is easy to see that

$$\sigma_{max}(n)\mu_{n,k}^{\text{mod}} \geq \sigma_{i,n,k}\mu_{n,k}^{\text{mod}} \qquad \forall n,k \tag{3.88}$$

since $\sigma_{max}(n) \geq \sigma_{i,n,k},$. Recall that the function $f_{i,j} \stackrel{\triangle}{=} \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i}$ is a standard function. From (3.88) and the monotonicity property of standard functions, we have

$$\frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{\text{mod}}} \geq \frac{\gamma_{i,j}}{\sigma_{i,i,j}\bar{m}_i^{\text{mod1}}} \tag{3.89}$$

where $\bar{m}_i^{\text{mod}}$ is given in (3.87) and $\bar{m}_i^{\text{mod1}}$ satisfy

$$\begin{aligned}\frac{1}{\bar{m}_i^{\text{mod1}}} = \frac{1}{N_t}\frac{\sigma_{i,i,k}\mu_{i,k}^{\text{mod}}}{1+\sigma_{i,i,k}\mu_{i,k}^{\text{mod}}\bar{m}_i^{\text{mod1}}} \\ + \frac{1}{N_t}\sum_{\substack{n=1\\n\neq i}}^{N}\sum_{k=1}^{K}\frac{\sigma_{i,n,k}\mu_{n,k}^{\text{mod}}}{1+\sigma_{i,n,k}\mu_{n,k}^{\text{mod}}\bar{m}_i^{\text{mod1}}} + 1.\end{aligned} \tag{3.90}$$

Notice that superscript corresponding to mod1 represents the original system with an uplink power allocation of $\boldsymbol{\mu}^{\text{mod}}$. Rearranging (3.89) we have, $\sigma_{i,i,j}\bar{m}_i^{\text{mod1}} \geq \sigma_{i,i,j}\bar{m}_i^{\text{mod}}$. Multiplying by $\mu_{i,j}^{\text{mod}}$, it follows that

$$\sigma_{i,i,j}\mu_{i,j}^{\text{mod}}\bar{m}_i^{\text{mod1}} \geq \sigma_{i,i,j}\mu_{i,j}^{\text{mod}}\bar{m}_i^{\text{mod}}. \tag{3.91}$$

The left hand side of (3.91) is the asymptotic uplink SINR of the original system with uplink power allocation $\boldsymbol{\mu}^{\text{mod}}$. From (3.86), the right hand side of equation (3.91) is equal to $\gamma_{i,j}$. Therefore,

$$\Lambda_{i,j}^{\text{org}}(\boldsymbol{\mu}^{\text{mod}}) \geq \gamma_{i,j}.$$

Hence we conclude that $\boldsymbol{\mu}^{\text{mod}}$ is a feasible power allocation for the original system.

3.7.5 Convergence proof of the Downlink SINR

In order to prove the convergence of the downlink SINR, we proceed as follows. Recall from Appendix B that the terms of the numerator and the denominator in the expression of $\Gamma_{i,j}(\mu_{i,j})$ converge to their deterministic equivalents. Also for any target SINR which satisfy feasibility conditions in (3.24), the uplink problem is feasible. Additionally, from Lemma 2, feasibility of the uplink problem implies the feasibility of the downlink. Further from Lemma 2, it also implies that the matrix $\mathbf{I} - \boldsymbol{\Gamma}\boldsymbol{\Delta}$ is invertible and hence there is a unique solution to the set of linear equations determining $\bar{\boldsymbol{\delta}}$. Once this is established, it follows that there is a unique mapping between the achieved SINR in the uplink and the downlink. Hence, the achieved SINR in the downlink must also converge to $\gamma_{i,j}$.

Chapter 4

Energy Efficient Design in MIMO Multi-cell Systems with Time Average QoS Constraints

Abstract: *In this work, we address the issue of energy efficient design in a MIMO multi-cell network consisting of N cells, N_t antennas per BS and K UTs per cell. Under this set up, we address the following question: given certain time average QoS targets for the users, what is the minimum energy expenditure with which they can be met? Time average QoS constraints can lead to greater energy savings as compared to instantaneous QoS constraints since it provides the flexibility to dynamically allocate resources over the fading channel states. We formulate the problem as a stochastic optimization problem whose solution is the design of the downlink beamforming vectors during each time slot. We first characterize the set of time average QoS targets which is achievable by some feasible control policy. We then use the technique of* virtual queue *to model the time average QoS constraints and convert the problem into a queue stabilization problem while minimizing the time average energy expenditure. We solve this problem using the approach of Lyapunov optimization and characterize its performance. Interestingly, our solution leads to a decentralized design in which the BSs only have to exchange limited side information. We also characterize the performance of our algorithm with delayed information exchange among the BSs.*

4.1 Introduction

Energy efficiency is becoming an important concern in the design of future wireless networks both from environmental and economical point of view. The exponential rise in the demand for wireless data services is expected to result in a significant increase in ICT related carbon emissions [16, 65]. Also, it has been observed that the major share in the costs of running a network are due to the energy bill [17] and significant savings can be achieved by reducing the network energy consumption. The above factors motivate us to consider the question of how future cellular networks can satisfy the QoS demands of users in an energy efficient manner.

There have been a number of works which have focused on the design of energy efficient communication systems. The focus of this work is primarily on energy efficiency in MIMO systems. In this context, relevant works include [66] which treats energy efficiency in a single user MIMO system from an information theoretic point of view. It considers optimizing the precoding matrix with the objective of maximizing the ratio of data rate to the consumed power. The effect of circuit power in addition to the transmission power is considered in [67] which optimizes the modulation and transmission strategy to minimize the total energy expenditure. [68] proposes switching between single user (SU) MIMO and multi user (MU) MIMO to achieve energy efficiency in the uplink. Other relevant works focusing on energy efficiency in MIMO systems include [69, 70, 71].

In this work, we consider the problem of energy efficient transmission techniques in a MIMO multi-cell scenario. We consider the problem of minimizing the time average energy expenditure subject to time average QoS constraints rather than instantaneous QoS constraints. The QoS metric we refer to in this work is the difference between the time average useful signal power and the time average interference signal power which is constrained to be greater than the target value. Our motivation to consider the time average QoS constraint comes from the fact that it provides the flexibility to dynamically allocate resources over the fading channel states as compared to instantaneous QoS constraints. It has been shown in many works that time average QoS constraints lead to better system performance as compared to the instantaneous (peak) constraints [72, 73, 74]. In terms of energy savings, time average QoS constraint can lead to better performance, due to the fact that the transmissions can delayed until favorable channel conditions are seen, thus minimizing the energy expenditure. The concept has also been exploited in the context of *energy-delay trade offs* [75, 76].

From a practical application point of view, our motivation to consider time average QoS constraint mainly comes from the fact that applications such as data services (such as file sharing, video download) allow some latitude in terms of delay tolerance.

We formulate our problem as a stochastic optimization problem and propose a solution based on the technique of Lyapunov optimization [13, 77]. Traditional approaches such as dynamic programming to solve stochastic optimization problems suffer from very high complexity and the fact that they require a-priori knowledge of the statistics of all the random processes in the system. In contrast, the technique of Lyapunov optimization is a powerful tool that provides simple online solutions based only on the current knowledge of the system state. Of particular relevance is the work of [78] which deals with the problem of power allocation to support the incoming traffic while minimizing the average power expenditure in wireless networks. The authors formulate a dynamic power allocation strategy for minimizing energy expenditure in a time varying wireless network using the technique of Lyapunov optimization. However, the interference model considered in [78] is a simple conflict graph based model in which if two transmitters are within the interference range of each other, only one of them can transmit. Therefore, the rate achieved depends only on the power allocated to the link which is scheduled (and hence the signal to noise ratio (SNR) of the link). However, advances in physical layer techniques like multiple antenna technologies allow multiple users to coexist over the same resource block. By using the technique of Lyapunov optimization, we obtain a series of optimization problems which must be solved during every time slot. In this work, we reformulate the optimization problem as a SDP problem and provide a method to compute the optimal beamforming vectors during each time slot.

Our contributions in this work are as follows.

- We first formulate the *feasible QoS region*, i.e. the set of target QoS constraints that is achievable by some control policy.
- We model the time average constraint as a *virtual queue* and transform the problem into a queue stabilization problem while minimizing the average energy expenditure. We then use the technique of Lyapunov optimization [13, 77] to formulate a dynamic control strategy that satisfies the time average QoS targets. Our algorithm leads to a decentralized design in which the BSs can formulate the beamforming vectors using only the local CSI. The BSs would only have to exchange virtual queue-length information among themselves.

- We provide the performance bounds for our decentralized algorithm and show that its performance in terms of the energy expenditure can be made arbitrarily close to the energy expenditure of the optimal policy.
- We further characterize the performance of our algorithm in the case with delayed queue-length information exchange among the BSs. We show that any time average QoS target which can be achieved under perfect queue-length information exchange is also achievable under the delayed queue-length information exchange. The delays however affects the performance bound in terms of the energy expenditure upto a constant value.

4.2 System Model

We consider a MIMO multi-cell scenario consisting N cells and K UTs per cell. The UTs in each cell are served by their respective BSs which are equipped with N_t antennas. The UTs have a single antenna each. The notation $\mathrm{UT}_{i,j}$ denotes the j-th UT present in the i-th cell. The BS of each cell serves only the UTs present in its cell. We consider a discrete-time block-fading channel model where the channel remains constant for a given coherence interval and then changes independently from one block to the other. We index the time slots by t. We denote the channel vector from the BS_i to the $\mathrm{UT}_{j,k}$ during the time slot t by $\mathbf{h}_{i,j,k}[t] \in \mathbb{C}^{N_t}$. We define the channel matrix $\mathbf{H}[t]$ given

$$\begin{aligned}\mathbf{H}_{i,j}[t] &= [\mathbf{h}_{i,j,1}[t], \dots, \mathbf{h}_{i,j,K}[t]] \\ \mathbf{H}_{i}[t] &= [\mathbf{H}_{i,1}[t], \dots, \mathbf{H}_{i,N}[t]] \\ \mathbf{H}[t] &= [\mathbf{H}_{i}[t], \dots, \mathbf{H}_{N}[t]]\,.\end{aligned}$$

The channel process $\{\mathbf{H}[t], t = 0, 1, 2, \dots\}$ is assumed to be an independent and identically distributed (i.i.d) discrete time stationary ergodic random process which takes values from the finite state space $\{\mathcal{H}_1, \dots, \mathcal{H}_L\}$. We denote the distribution

$$\Pr(\mathbf{H}[t] = \mathcal{H}_l) = \pi_l, \qquad l = 1, \dots, L$$

where

$$\sum_l \pi_l = 1.$$

Let us denote the beamforming vector corresponding to $\mathrm{UT}_{i,j}$ during slot t by $\mathbf{w}_{i,j}[t] \in \mathbb{C}^{N_t}$. We also denote

$$\mathbf{W}_i[t] = [\mathbf{w}_{i,1}[t], \ldots, \mathbf{w}_{i,K}[t]]$$
$$\mathbf{W}[t] = [\mathbf{W}_i[t], \ldots, \mathbf{W}_N[t]]\,.$$

The signal received by $\mathrm{UT}_{i,j}$ during time t is given by

$$y_{i,j}[t] = \mathbf{h}_{i,i,j}^H[t]\mathbf{w}_{i,j}[t]x_{i,j}[t] + \sum_{\substack{(n,k)\\ \neq(i,j)}} \mathbf{h}_{n,i,j}^H[t]\mathbf{w}_{n,k}[t]x_{n,k}[t] + z_{i,j}[t] \tag{4.1}$$

where $x_{i,j}[t] \in \mathbb{C}$ represents the information signal for the $\mathrm{UT}_{i,j}$ during the time slot t and $z_{i,j} \sim \mathcal{CN}(0, N_0)$ is the corresponding additive white Gaussian complex noise. The QoS metric which we denote by $\gamma_{i,j}[t]$ is

$$\gamma_{i,j}[t] = |\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2 - \sum_{\substack{(n,k)\\ \neq(i,j)}} |\mathbf{w}_{n,k}^H[t]\mathbf{h}_{n,i,j}[t]|^2. \tag{4.2}$$

We denote the time average QoS metric by

$$\bar{\gamma}_{i,j} = \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\gamma_{i,j}[t]\right]. \tag{4.3}$$

The QoS metric chosen in this work represents the difference between time average useful signal power and the time average interference signal power. The QoS metric in our work is in similar spirit with metrics such as interference temperature control [79, 80] (in which the interference, peak or average value is constrained to below a certain threshold) or the signal preservation constraint [81] (in which the interference power is minimized subject to preserving the desired component of the received signal). In this work, we consider the time average QoS constraints as it can provide more flexibility as compared to instantaneous constraints in terms of dynamically allocating resources over fading channel states, thus leading to better system performance.

The transmission power by each BS $P_i[t]$ depends on the beamforming vector during the time slot t which can be given as

$$P_i[t] = \sum_{j=1}^{K} \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t] \qquad i = 1, \ldots, N. \tag{4.4}$$

The optimization problem to minimize the average energy expenditure subject

to time average QoS constraint can be formulated as

$$\min \quad \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\Big[\sum_{i=1}^{N} P_i[t]\Big] \tag{4.5}$$

$$s.t. \quad \bar{\gamma}_{i,j} \geq \lambda_{i,j}, \qquad \forall i,j \tag{4.6}$$

$$\sum_{j=1}^{K} \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t] \leq P_{\text{peak}} \qquad \forall i,t \tag{4.7}$$

where P_{peak} is the peak power at which the BSs can transmit.

The optimization problem in (4.5) is a stochastic optimization problem. The control action to be taken during each time slot is the formulation of the downlink beamforming vectors ($\mathbf{w}_{i,j}[t]\ \forall i,j$) during every time slot t. In particular, we search for the sequence of control actions which result in minimum time average energy expenditure while satisfying the time average QoS constraints. Let us denote the time average energy expenditure by

$$\bar{P} = \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\sum_{i=1}^{N} P_i[t]\right]. \tag{4.8}$$

Let P_{inf} be the infimum time average power $\bar{P}$ incurred while achieving the time average QoS targets over all possible sequence of control actions.

The stochastic optimization problem (4.5) can be solved using techniques such as dynamic programming. However, these are computationally complex and require the knowledge of the statistics of the associated random processes. Instead in this work, we use the approach of Lyapunov optimization [13] to solve the problem. The Lyapunov optimization method leads to a simple online solution which requires only the knowledge of the current channel conditions.

4.3 Achievable QoS Region

Before we solve (4.5), we will first characterize the set of all the time average QoS targets which can be satisfied by some control policy. Problem (4.5) is solvable only if the time average QoS constraints lie within this region. Let us define the matrices

$$\mathbf{W}_i = [\mathbf{w}_{i,1}, \ldots, \mathbf{w}_{i,K}] \in \mathbb{C}^{N_t \times K} \tag{4.9}$$

$$\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_N] \in \mathbb{C}^{N_t \times NK}. \tag{4.10}$$

We denote the collection of all feasible beamforming vectors $\mathbf{W}$ by the set $\mathcal{W}$ which can be characterized as follows,

$$\mathcal{W} = \{\mathbf{W} \ : \ \mathbf{w}_{i,j} \in \mathbb{C}^{N_t} \ \forall i,j \\ \text{s.t.} \ \sum_j \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \leq P_{\max}, \ \forall i\}. \tag{4.11}$$

The QoS associated with a particular beamforming policy $\mathbf{W}$ when the channel is in state l is

$$\gamma_{i,j}(l, \mathbf{W}) = |\mathbf{h}_{i,i,j}^H \mathbf{w}_{i,j}|^2 - \sum_{\substack{(n,k) \\ \neq (i,j)}} |\mathbf{h}_{n,i,j}^H \mathbf{w}_{n,k}|^2. \tag{4.12}$$

We also define the vectors

$$\boldsymbol{\gamma}_i(l, \mathbf{W}) = [\gamma_{i,1}, \ldots, \gamma_{i,K}] \in \mathbb{R}^K \tag{4.13}$$

$$\boldsymbol{\gamma}(l, \mathbf{W}) = [\gamma_1, \ldots, \gamma_N] \in \mathbb{R}^{NK}. \tag{4.14}$$

The *achievable QoS region* is then given by

$$\boldsymbol{\Gamma} = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^{NK} : \boldsymbol{\lambda} \leq \sum_l \pi_l \mathcal{CH}\big(\boldsymbol{\gamma}(l, \mathbf{W}) \big| \mathbf{W} \in \mathcal{W}\big) \right\} \tag{4.15}$$

where $\mathcal{CH}(\mathbf{A})$ represents the convex hull of the set $\mathbf{A}$.

For any target QoS lying inside the stability region ($\boldsymbol{\lambda} \in \boldsymbol{\Gamma}$), there exists a stationary randomized policy ([82], Theorem 4.5) that can achieve the target QoS.

4.4 Energy Efficient Decentralized Beamforming Design

In this work, in order to model the time average QoS constraint, we use the concept of *virtual queue* [82]. The virtual queue associated with the time average constraint $\bar{\gamma}_{i,j} \geq \lambda_{i,j}$ evolves in the following manner,

$$Q_{i,j}[t+1] = \max\left(Q_{i,j}[t] - \mu_{i,j}[t], 0\right) + A_{i,j}[t] \tag{4.16}$$

where $A_{i,j}[t]$ denotes the arrival process

$$A_{i,j}[t] = \sum_{\substack{(n,k) \\ \neq (i,j)}} |\mathbf{w}_{n,k}^H[t] \mathbf{h}_{n,i,j}[t]|^2 + \lambda_{i,j} \tag{4.17}$$

and $\mu_{i,j}[t]$ denotes the departure process

$$\mu_{i,j}[t] = |\mathbf{w}_{i,j}^H[t] \mathbf{h}_{i,i,j}[t]|^2. \tag{4.18}$$

The notion of strong stability of the virtual queue is given as follows,

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\big[\sum_{i,j} Q_{i,j}[t]\big] < \infty. \tag{4.19}$$

Ensuring the strong stability of the virtual queue implies that the lim sup of time average of the arrival process is less than or equal to the service process [82][1], i.e.

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[A_{i,j}[t] - \mu_{i,j}[t]] \leq 0 \qquad \forall i,j. \tag{4.20}$$

In other words, the constraint (4.6) is satisfied. Thus, we reformulate the original problem into a problem of stabilizing the virtual queue while minimizing the time average energy expenditure as

$$\begin{aligned} \min \quad & \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\Big[\sum_{i,j} \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\Big] \\ s.t. \quad & \text{Virtual Queue Stability} \\ & \sum_{j} \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t] \leq P_{\text{peak}} \qquad \forall i,t. \end{aligned} \tag{4.21}$$

In order to solve the above problem, we use the technique of Lyapunov optimization [13] which allows us to consider the joint problem of stabilizing the queue and performance optimization.

To this end, we define the quadratic Lyapunov function $V : \mathbb{R}^N \to \mathbb{R}$ as follows:

$$V(\mathbf{Q}[t]) = \frac{1}{2} \sum_{i,j} (Q_{i,j}[t])^2. \tag{4.22}$$

The Lyapunov function is a scalar measure of the aggregate queue-lengths in the system. We define the one-step conditional Lyapunov drift as

$$\Delta(\mathbf{Q}[t]) = \mathbb{E}\left[V(\mathbf{Q}[t+1])) - V(\mathbf{Q}[t])\Big|\mathbf{Q}[t]\right] \tag{4.23}$$

where the expectation is with respect to the random channel states and the (possibly random) control actions made in reaction to these channel states. We now define the following result which will be helpful in formulating the online beamforming algorithm.

Theorem 7. *(Lyapunov Optimization [13]) If there exist constants $B > 0$, $\epsilon > 0$, $V > 0$ such that for all time slots t we have,*

$$\Delta(\mathbf{Q}[t]) + V\mathbb{E}\Big[\sum_i P_i[t]\Big|\mathbf{Q}(t)\Big] \leq B - \epsilon \sum_{i,j} Q_{i,j}[t] + VP_{\text{inf}} \tag{4.24}$$

[1] In general mean rate stability is sufficient to ensure this condition. Here we use a stronger notion of stability.

then the system is stable and time average backlog satisfies:

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\sum_{i,j}Q_{i,j}[t]\right]\leq\frac{B+VNKP_{peak}}{\epsilon} \tag{4.25}$$

and

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\Big[\sum_{i}P_i[t]\Big]\leq P_{\text{inf}}+B/V. \tag{4.26}$$

Theorem 7 suggests that if the control action (beamforming design) is formulated such that the condition of (4.24) is satisfied for every time slot t, then the virtual queue is strongly stable (4.25). Further, from (4.26) the time average energy expenditure can be made arbitrarily close to the optimal value of this problem by increasing the value of V.

We will now examine the Lyapunov drift corresponding to the evolution of the virtual queue $Q_{i,j}$.

Proposition 3. *For the virtual queue which evolves according to* (4.16)*, the Lyapunov drift follows the following condition.*

$$\begin{aligned}\Delta(\mathbf{Q}[t])+V\mathbb{E}\big[\sum_{i,j}\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\big|\mathbf{Q}[t]\big]\leq C_1\\ +\sum_{i,j}Q_{i,j}[t]\lambda_{i,j}-\sum_{i,j}\mathbb{E}\Big[Q_{i,j}[t]\Big(|\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2-\sum_{\substack{(n,k)\\\neq(i,j)}}|\mathbf{w}_{n,k}^H[t]\mathbf{h}_{n,i,j}[t]|^2\Big)\\ -V\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\Big|\mathbf{Q}[t]\Big]\qquad\forall t\end{aligned} \tag{4.27}$$

where

$$C_1=\frac{1}{2}\sum_{i,j}\mathbb{E}\Big[(A_{i,j}^{\max}[t])^2+(\mu_{i,j}^{\max}[t])^2\Big]<\infty \tag{4.28}$$

and

$$A_{\max}[t]=NP_{peak}\max_n|\mathbf{h}_{n,i,j}[t]|^2+\lambda_{i,j} \tag{4.29}$$

$$\mu_{\max}[t]=P_{peak}|\mathbf{h}_{i,i,j}[t]|^2. \tag{4.30}$$

The proof follows from the steps in Subsection 4.9.1, part I. According to the theory of stochastic network optimization, a good method to choose the beamforming vector is to minimize the bound obtained in (4.27). This implies that the beamforming vector should be chosen in the following manner:

$$\begin{aligned}\mathbf{w}_{i,j}[t]\quad\in\arg\max_{\mathbf{w}\in\mathcal{W}}\quad&\sum_{i,j}\mathbb{E}_{\mathbf{H}}\Big[Q_{i,j}[t]|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}[t]|^2\\ &-Q_{i,j}[t]\sum_{\substack{(n,k)\\\neq(i,j)}}|\mathbf{w}_{n,k}^H\mathbf{h}_{n,i,j}[t]|^2-V\mathbf{w}_{i,j}^H\mathbf{w}_{i,j}\Big|\mathbf{Q}[t]\Big]\end{aligned} \tag{4.31}$$

where $\mathbb{E}_{\mathbf{H}}$ indicates that the expectation is with respect to the random channel realization. We will assume that the BSs have the perfect knowledge of CSI of all its downlink channels ($\mathbf{h}_{i,n,k}\ \forall n,k$)and propose a method to solve (4.31) using the approach of SDP. Interestingly, our problem approach also leads to a decentralized solution in the multi-cell scenario. Further we will theoretically examine some properties of this algorithm and analyze its performance.

4.4.1 Perfect CSI Case

The BS has perfect knowledge of the channel $\mathbf{H}[t]$. With the perfect knowledge of CSI, the optimization problem (4.31) becomes

$$\begin{aligned}\mathbf{w}_{i,j}[t] \quad \in \arg\max_{\mathbf{w}\in\mathcal{W}} \quad & \sum_{i,j} \mathbb{E}_{\mathbf{H}}\Big[Q_{i,j}[t]|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}[t]|^2 \\ & -Q_{i,j}[t]\sum_{\substack{(n,k)\\ \neq(i,j)}}|\mathbf{w}_{n,k}^H\mathbf{h}_{n,i,j}[t]|^2 - V\mathbf{w}_{i,j}^H\mathbf{w}_{i,j}\Big|\mathbf{Q}[t],\mathbf{H}[t]\Big]\end{aligned} \tag{4.32}$$

which reduces to greedily minimizing the term inside the expectation ($\mathbb{E}[f(Y)|Y] = f(Y)$). Therefore, we remove the expectation and solve the following optimization problem (we drop the time index t),

$$\begin{aligned}\max_{\mathbf{w}\in\mathbb{C}^{N_t}} \quad & \sum_{i,j}\Big[Q_{i,j}|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}|^2 \\ & -Q_{i,j}\sum_{\substack{(n,k)\\ \neq(i,j)}}|\mathbf{w}_{n,k}^H\mathbf{h}_{n,i,j}|^2 - V\mathbf{w}_{i,j}^H\mathbf{w}_{i,j}\Big] \\ s.t. \quad & \sum_j \mathbf{w}_{i,j}^H\mathbf{w}_{i,j} \le P_{\text{peak}} \qquad \forall i.\end{aligned} \tag{4.33}$$

The objective function of the optimization problem in (4.33) can be rearranged as

$$\begin{aligned}\max_{\mathbf{w}\in\mathbb{C}^{N_t}} \quad & \sum_{i,j}\Big[Q_{i,j}|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}|^2 \\ & -\sum_{\substack{(n,k)\\ \neq(i,j)}}Q_{n,k}|\mathbf{w}_{i,j}^H\mathbf{h}_{i,n,k}|^2 - V\mathbf{w}_{i,j}^H\mathbf{w}_{i,j}\Big] \\ s.t. \quad & \sum_j \mathbf{w}_{i,j}^H\mathbf{w}_{i,j} \le P_{\text{peak}}. \qquad \forall i\end{aligned} \tag{4.34}$$

Note that we can write the quadratic terms in the objective function of (4.34) as

$$\begin{aligned}|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}|^2 &= \mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}\mathbf{h}_{i,i,j}^H\mathbf{w}_{i,j} \\ &= \mathbf{w}_{i,j}^H\mathbf{H}_{i,i,j}\mathbf{w}_{i,j}\end{aligned} \tag{4.35}$$

where the matrix $\mathbf{H}_{i,i,j} = \mathbf{h}_{i,i,j}\mathbf{h}_{i,i,j}^H$. Similarly we have,

$$|\mathbf{w}_{i,j}^H \mathbf{h}_{i,n,k}|^2 = \mathbf{w}_{i,j}^H \mathbf{H}_{i,n,k} \mathbf{w}_{i,j}. \tag{4.36}$$

Using (4.35) and (4.36), the optimization problem of (4.34) can be written as,

$$\begin{aligned} \max_{\mathbf{w}} \quad & \sum_{i,j} \mathbf{w}_{i,j}^H \mathbf{A}_{i,j} \mathbf{w}_{i,j} \\ s.t. \quad & \sum_{j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \leq P_{\text{peak}} \qquad \forall i \end{aligned} \tag{4.37}$$

where the matrix

$$\mathbf{A}_{i,j} = Q_{i,j}\mathbf{H}_{i,i,j} - \sum_{\substack{(n,k) \\ \neq (i,j)}} Q_{n,k}\mathbf{H}_{i,n,k} - V\mathbf{I}. \tag{4.38}$$

Note that the optimization problem in (4.37) is in separable form, each BS i can solve the optimization problem given by

$$\begin{aligned} \max_{\mathbf{w}} \quad & \sum_{j} \mathbf{w}_{i,j}^H \mathbf{A}_{i,j} \mathbf{w}_{i,j} \\ s.t. \quad & \sum_{j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \leq P_{\text{peak}}. \end{aligned} \tag{4.39}$$

In order to solve (4.39), we further rearrange the quadratic term as

$$\begin{aligned} \mathbf{w}_{i,j}^H \mathbf{A}_{i,j} \mathbf{w}_{i,j} &= \text{tr}(\mathbf{w}_{i,j}^H \mathbf{A}_{i,j} \mathbf{w}_{i,j}) \\ &= \text{tr}\left(\mathbf{A}_{i,j} \mathbf{w}_{i,j} \mathbf{w}_{i,j}^H\right) \\ &= \text{tr}\left(\mathbf{A}_{i,j} \mathbf{W}_{i,j}\right) \end{aligned} \tag{4.40}$$

where the matrix $\mathbf{W}_{i,j} = \mathbf{w}_{i,j}\mathbf{w}_{i,j}^H$. (4.39) can now be rewritten as

$$\begin{aligned} \max_{\mathbf{W}} \quad & \sum_{j} \text{tr}\left(\mathbf{A}_{i,j} \mathbf{W}_{i,j}\right) \\ s.t. \quad & \sum_{j} \text{tr}(\mathbf{W}_{i,j}) \leq P_{\text{peak}} \\ & \text{rank}(\mathbf{W}_{i,j}) = 1 \qquad \forall j. \end{aligned} \tag{4.41}$$

In what follows, we will consider the SDP relaxation of (4.41) and remove the rank constraint on the $\mathbf{W}_{i,j}$ matrices

$$\begin{aligned} \max_{\mathbf{W}} \quad & \sum_{j} \text{tr}\left(\mathbf{A}_{i,j} \mathbf{W}_{i,j}\right) \\ s.t. \quad & \sum_{j} \text{tr}(\mathbf{W}_{i,j}) \leq P_{\text{peak}}. \end{aligned} \tag{4.42}$$

Problem (4.42) is in standard SDP form and can be solved using the optimization package SEDUMI [83]. Also, from the result of [84], it is easy to see that the solution to (4.42) is always rank 1. Therefore, the solution to the SDP problem directly gives optimal beamforming vectors in our case. SDP problems can be efficiently solved using interior point methods, at a complexity cost of at most $O(N + N_t^2)^{3.5}$. Later on in this section, we will provide an algorithm to solve (4.42) with a lower complexity.

Observe that in order to formulate the matrix $\mathbf{A}_{i,j}$, the BSs only require the local CSI ($\mathbf{h}_{i,n,k}\ \forall n, k$). The BSs would only have to exchange the queue-lengths among themselves. Therefore, our formulation naturally leads to a decentralized solution.

We will hereby address the algorithm corresponding to solving (4.39) as the decentralized beamforming (DBF) algorithm. Further, we will denote the solution corresponding to (4.39) by the superscript "opt". We will now provide some theoretical analysis into the performance of the DBF algorithm.

Proposition 4. *Under the DBF strategy, during every time slot t, it is optimal for BS to serve at most one UT per cell. Further, during a given time slot t, if the BS is serving a UT, then it is optimal to serve it at its peak power.*

Proposition 4 can be argued as follows. Recall the optimization problem in (4.37). Let us decompose the beamforming vector into power allocation and direction vector as follows.

$$\begin{aligned} ||\mathbf{w}_{i,j}||^2 &= P_{i,j} \\ \mathbf{d}_{i,j} &= \frac{\mathbf{w}_{i,j}}{||\mathbf{w}_{i,j}||}. \end{aligned} \tag{4.43}$$

Using (4.43), optimization problem in (4.37) can be written as

$$\begin{aligned} \max_{P_{i,j}, \mathbf{d}_{i,j}} \quad & \sum_j P_{i,j} \mathbf{d}_{i,j}^H \mathbf{A}_{i,j} \mathbf{d}_{i,j} \\ s.t. \quad & \sum_j P_{i,j} \leq P_{\text{peak}} \\ & ||\mathbf{d}_{i,j}|| = 1 \qquad \forall i, j. \end{aligned} \tag{4.44}$$

Equation (4.43) decouples the direction vectors $\mathbf{d}_{i,j}$. It can be easily verified that the optimization problem in (4.44) can be solved as

$$\begin{aligned} \max_{P_{i,j}} \quad & \sum_j P_{i,j} \left(\max_{||\mathbf{d}_{i,j}||=1} \mathbf{d}_{i,j}^H \mathbf{A}_{i,j} \mathbf{d}_{i,j} \right) \\ s.t. \quad & \sum_j P_{i,j} \leq P_{\text{peak}}. \end{aligned} \tag{4.45}$$

Note that from Rayleigh-Ritz theorem (Chapter 2, Theorem 1), it follows that

$$\max_{||\mathbf{d}_{i,j}||=1} \mathbf{d}_{i,j}^H \mathbf{A}_{i,j} \mathbf{d}_{i,j} = \lambda^{\max}(\mathbf{A}_{i,j}) \tag{4.46}$$

where $\lambda^{\max}(\mathbf{A}_{i,j})$ is the maximum eigen value of the matrix $\mathbf{A}_{i,j}$. Using (4.46) in (4.45) yields

$$\begin{aligned} \max_{P_{i,j}} \quad & \sum_j P_{i,j} \lambda^{\max}(\mathbf{A}_{i,j}) \\ s.t. \quad & \sum_j P_{i,j} \leq P_{\text{peak}}. \end{aligned} \tag{4.47}$$

Let us define

$$j^* = \arg\max_j \lambda^{\max}(\mathbf{A}_{i,j}).$$

Therefore from (4.47), it is straightforward to see that,

$$P_{i,j}^{\text{opt}} = \begin{cases} P_{\text{peak}} & \text{if } j = j^* \text{and } \lambda^{\max}(\mathbf{A}_{i,j^*}) > 0 \\ 0 & \text{else.} \end{cases} \tag{4.48}$$

Proposition 4 implies that at most one UT can be active per cell during each time slot. During a given time slot t, in every cell, there can be either one active UT in the cell which case the BS transmits at its peak power ($P_i^{\text{opt}}[t] = P_{\text{peak}}$) or there are no active UTs in the cell and hence the BS does not transmit ($P_i^{\text{opt}}[t] = 0$). Also, we can conclude that

$$\begin{aligned} &\sum_j \text{tr}(\mathbf{A}_{i,j}[t] \mathbf{W}_{i,j}^{\text{opt}}[t]) \\ &\quad = P_{\text{peak}} \lambda^{\max}(\mathbf{A}_{i,j^*}) \mathbf{1}_{\lambda^{\max}(\mathbf{A}_{i,j^*})>0}. \end{aligned} \tag{4.49}$$

Intuition:

Proposition 4 provides us an intuitive way to understand the working of the DBF algorithm. Consider a two cell scenario as shown in Figure 4.1 consisting of one UT per cell. Also lets assume that each BS has only a single antenna. Therefore, the DBF algorithm now corresponds to a power allocation problem. Let us denote the channel from BS_i to UT_i by h_{ii} and from BS_i to UT_j by $h_{ij},\ j \neq i$ and the queue-length of UT_i by Q_i. Let us define the terms,

$$\begin{aligned} A_1[t] &= Q_1[t]|h_{11}[t]|^2 - Q_2[t]|h_{12}[t]|^2 - V \\ A_2[t] &= Q_2[t]|h_{11}[t]|^2 - Q_1[t]|h_{12}[t]|^2 - V. \end{aligned}$$

The optimization problem corresponding to DBF algorithm now reduces to

$$\max_{\substack{0 \leq P1 \leq P_{\text{peak}} \\ 0 \leq P2 \leq P_{\text{peak}}}} A_1[t]P_1 + A_2[t]P_2.$$

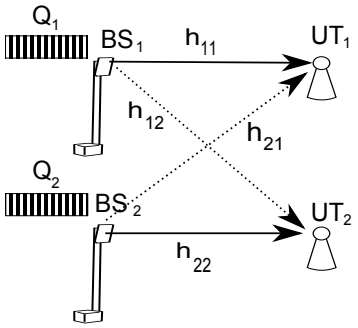

Figure 4.1: Two cell scenario

Then, the optimal power allocation policy is given

$$P_i[t] = \begin{cases} P_{\text{peak}} & \text{if } A_i[t] > 0 \\ 0 & \text{else.} \end{cases} \tag{4.50}$$

for $i = 1, 2$. The result can be interpreted as follows. BS_i decides to transmit if the weighted sum of the useful signal power to UT_i exceeds the weighted sum of the interference signal power it causes to UT_j. The weights in our case are the virtual queue-length. During a given time slot, the BS may decide to serve the UT due to the following factors,

- High value of the queue-length $Q_i[t]$ which implies that the queue of UT_i has a high backlog and hence needs to be served.
- The channel corresponding to the useful signal $h_{ii},\ i = 1, 2$ is good.
- The interference caused to the other UTs is low.

The factor V is an indicator of how aggressively the BS decides has to transmit. Higher value of V implies less frequent transmissions and hence less energy expenditure. However, it also implies that the virtual queue-lengths have to grow to a higher value before they are served, hence resulting in a higher average queue-length.

Proposition 4 also provides us an easier method to solve the optimization problem (4.33). It can be seen that the optimal beamforming vector during each time slot is the eigen vector corresponding to the maximum eigen value ($\lambda^{\max}(\mathbf{A}_{i,j^*})$.) Therefore,

$$\mathbf{w}_{i,j^*}^{\text{opt}} = \left(P_{\text{peak}} \lambda^{\max}(\mathbf{A}_{i,j^*}) \mathbf{1}_{\lambda^{\max}(\mathbf{A}_{i,j^*})>0} \right) \mathbf{x}_{\lambda^{\max}(\mathbf{A}_{i,j^*})} \tag{4.51}$$

where $\mathbf{x}_{\lambda^{\max}(\mathbf{A}_{i,j^*})}$ is the eigen vector corresponding to the maximum eigen value of matrix $\mathbf{A}_{i,j^*}$ and

$$\mathbf{w}_{i,j}^{\text{opt}} = \mathbf{0} \qquad j^* \neq j.$$

At this stage, we would like to point out that the complexity of eigenvalue decomposition function is $O(N_t^3)$ which is lower than the complexity of the SDP based solution. Therefore, (4.51) offers an easier method to calculate the optimal beamforming vectors with lesser computational complexity.

We will now proceed to provide performance bounds of the DBF algorithm in terms of average power expenditure and average backlog.

Proposition 5. *For any target SINR lying strictly inside the feasible QoS region, the DBF algorithm yields the following performance bounds. The virtual queue is strongly stable and for any $V > 0$, the time average queue-length satisfies*

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\sum_{i,j} Q_{i,j}^{opt}[t]\right] \leq \frac{C_1 + VNKP_{peak}}{\epsilon} \tag{4.52}$$

and the time average energy expenditure yields the following bounds,

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\sum_{i=1}^{N} P_i^{opt}[t]\right] \leq P_{\text{inf}} + \frac{C_1}{V}. \tag{4.53}$$

where C_1 is defined in (4.28).

The proof is provided in Subsection 4.9.1, part II. The bound in (4.53) implies that the time average energy expended by the DBF algorithm can be made arbitrarily close to the minimum average power (over all possible sequence on control actions) by increasing the value of V to an arbitrarily an high value. This comes at the expense of increasing the average queue-length of the virtual queue. Intutively, a high value of the average queue-length implies that the number of time slots required to satisfy the time average constraints is higher (analogous to the concept of delay in real queues).

We will now proceed to introduce delays in the information exchange among the BSs. Recall that the only information the BSs would have to exchange in the DBF algorithm are the queue-lengths. In what follows, we will study the impact of delayed queue-length information exchange on the performance of the DBF algorithm.

4.5 Delayed Queue-length Information Exchange

Let us assume that a delay of $\tau < \infty$ time slots is incurred while the BSs exchange the queue-length information. Each BS i now has perfect queue-length information of its local queues ($Q_{i,j}[t]\ \forall j$) and the delayed queue-length information from the neighboring queues ($Q_{n,k}[t-\tau],\ \forall n \neq i,k$). Note that our set up can be easily generalized to introduce different delays $\tau_n, \forall n \neq i$ corresponding to the queue-length information from different BSs. However, in order to keep the notations simple, we restrict ourselves to uniform delays ($\tau,\ \forall n \neq i$).

We assume that the BSs treat the delayed queue-length as the true value of the queue-length. Every BS now solves the following optimization problem,

$$\begin{aligned} \max_{\mathbf{W}} \quad & \sum_j \mathrm{tr}\left(\mathbf{A}^{\tau}_{i,j}[t]\mathbf{W}_{i,j}\right) \\ s.t. \quad & \sum_j \mathrm{tr}(\mathbf{W}_{i,j}) \leq P_{\mathrm{peak}} \end{aligned} \tag{4.54}$$

where the matrix $\mathbf{A}^{\tau}_{i,j}[t]$ is given by,

$$\begin{aligned} \mathbf{A}^{\tau}_{i,j}[t] = & Q_{i,j}[t]\mathbf{H}_{i,i,j}(t) - \sum_{k\neq j} \lambda_{i,k} Q_{i,k}[t]\mathbf{H}_{i,i,k}[t] \\ & - \sum_{n\neq i,k} \lambda_{n,k} Q_{n,k}[t-\tau]\mathbf{H}_{i,n,k}[t] - V\mathbf{I}. \end{aligned} \tag{4.55}$$

Let us denote the solution corresponding to optimization problem (4.54) by $\mathbf{W}^{\mathrm{del}}[t]$. We will henceforth use the superscript "del" to denote parameters corresponding to the solution of with delayed queue-length information. Once again, following similar argument as Proposition 4, it can be shown that at most one UT can be active per cell. Let us define

$$j^*_{\tau} = \arg\max_j \lambda^{\max}(\mathbf{A}^{\tau}_{i,j}).$$

It is clear that like (4.47), the optimal power allocation policy is given by

$$P^{\mathrm{del}}_{i,j} = \begin{cases} P_{\mathrm{peak}} & \text{if } j = j^*_{\tau} \text{and } \lambda^{\max}(\mathbf{A}_{i,j^*_{\tau}}) > 0 \\ 0 & \text{else} \end{cases} \tag{4.56}$$

and therefore,

$$\begin{aligned} & \max_{\mathbf{W}_{i,j}} \sum_j \mathrm{tr}(\mathbf{A}^{\tau}_{i,j}[t]\mathbf{W}^{\mathrm{del}}_{i,j}[t]) \\ & \qquad = P_{\mathrm{peak}} \lambda^{\max}(\mathbf{A}_{i,j^*_{\tau}}) \mathbf{1}_{\lambda^{\max}(\mathbf{A}_{i,j^*_{\tau}})>0}. \end{aligned} \tag{4.57}$$

We will now theoretically examine the performance of the DBF algorithm with delayed queue-length information. We will first compare the performance of DBF algorithm with perfect queue-length exchange with that of delayed queue-length exchange during each time slot t in the following lemma.

Lemma 12. *There exists a* $0 \leq C_2 < \infty$ *independent of the current queue-length* $Q_{i,j}[t],\ \forall i,j$ *such that,*

$$\sum_{i,j} tr(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{opt}[t]) \leq \sum_{i,j} tr(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{del}[t]) + C_2\ \forall t. \tag{4.58}$$

The lemma is proved in Subsection 4.9.2, part I. The lemma states the performance of the DBF algorithm with delayed queue-length information exchange differs from that of DBF algorithm with perfect queue-length information exchange by a bounded constant. The key element in this lemma is the fact that the constant C_2 is independent of the current queue-lengths which will be helpful in proving the performance bounds for the DBF algorithm with delayed queue-length information exchange.

Theorem 8. *Following the DBF algorithm with delayed queue-length information exchange, the following performance can be obtained. The time average queue-length satisfies,*

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i,j} \mathbb{E}\left[Q_{i,j}^{del}[t]\right] \leq \frac{C_1 + C_2 + VNKP_{peak}}{\epsilon} \tag{4.59}$$

and the time average energy expenditure satisfies

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{N} \mathbb{E}\left[P_i^{del}[t]\right] \leq P_{\text{inf}} + \frac{C_1 + C_2}{V}. \tag{4.60}$$

where C_1 *is defined in* (4.28) *and* C_2 *defined from Lemma 12.*

Thereom 8 is proved in Subsection 4.9.2, part II. From Theorem 8, it can be seen that even with delayed queue-length information exchange, the queue is strongly stable. Therefore, any time average QoS target achievable under perfect queue-length information exchange is also achievable under the delayed queue-length information exchange. The delay only affects the performance bound of the average energy expenditure upto a finite constant.

4.6 The Case with Channel Estimation

In this section, we will show how the framework of Lyapunov optimization can be extended to the case when the BSs do not have perfect CSI and have to

estimate the channel. The channel $\mathbf{h}_{i,i,j}[t]$ at each time slot t can be written in terms of the estimate $\hat{\mathbf{h}}_{i,i,j}[t]$ and estimation error $\tilde{\mathbf{h}}_{i,i,j}[t]$ as

$$\mathbf{h}_{i,i,j}[t] = \hat{\mathbf{h}}_{i,i,j}[t] + \tilde{\mathbf{h}}_{i,i,j}[t] \qquad \forall i, j. \tag{4.61}$$

We consider a minimum mean square error (MMSE) kind of estimator, in which $\tilde{\mathbf{h}}_{i,i,j}[t]$ independent of the $\hat{\mathbf{h}}_{i,i,j}[t]$ and whose covariance is given by $\sigma_e^2\mathbf{I}$ where σ_e^2 is a quantity that depends on the channel training time [85]. In general, σ_e^2 decreases with the channel training time. In this work, we do not consider the issue of optimizing the channel training time. Instead, we assume that the channel training time (and hence the accuracy of the estimate) is pre-determined according to certain performance criteria and the estimate $\hat{\mathbf{h}}_{i,i,j}[t]$ is known at the transmitter.

We will first characterize the achievable QoS region in the case when the BS does not have the perfect CSI. The QoS associated with a particular beamforming policy $\mathbf{W}$ when the channel is in state l and an estimate $\hat{\mathbf{H}}$ is known at the transmitter is given by

$$\gamma_{i,j}^{\text{est}}(l, \mathbf{W}) = \mathbb{E}_{\mathbf{H}}\Big[|\mathbf{h}_{i,i,j}^H \mathbf{w}_{i,j}|^2 - \sum_{\substack{(n,k)\\ \neq(i,j)}} |\mathbf{h}_{n,i,j}^H \mathbf{w}_{n,k}|^2 \Big| \hat{\mathbf{H}}\Big]. \tag{4.62}$$

We also define the vectors

$$\boldsymbol{\gamma}_i^{\text{est}}(l, \mathbf{W}) = [\gamma_{i,1}^{\text{est}}, \ldots, \gamma_{i,K}^{\text{est}}] \in \mathbb{R}^K \tag{4.63}$$

$$\boldsymbol{\gamma}^{\text{est}}(l, \mathbf{W}) = [\gamma_1^{\text{est}}, \ldots, \gamma_N^{\text{est}}] \in \mathbb{R}^{NK}. \tag{4.64}$$

The *achievable QoS region* in the case with channel estimate is given by

$$\boldsymbol{\Gamma}^{\text{est}} = \Big\{\boldsymbol{\lambda}^{\text{est}} \in \mathbb{R}^{NK} : \boldsymbol{\lambda}^{\text{est}} \leq \sum_l \pi_l \mathcal{CH}\big(\boldsymbol{\gamma}^{\text{est}}(l, \mathbf{W}) \big| \mathbf{W} \in \mathcal{W}\big)\Big\}. \tag{4.65}$$

Next, we consider the problem of finding the optimal beamforming vectors given a channel estimate in order to minimize the long term energy expenditure. To this end, consider the optimization problem (4.31). When the BS has only the knowledge of the channel estimate, the optimization problem becomes the following,

$$\begin{aligned} \max_{\mathbf{w}} \quad & \sum_{i,j} \Big[Q_{i,j} \mathbb{E}_{\mathbf{H}}\Big[|\mathbf{w}_{i,j}^H \mathbf{h}_{i,i,j}|^2 \\ & -Q_{i,j} \sum_{\substack{(n,k)\\ \neq(i,j)}} |\mathbf{w}_{n,k}^H \mathbf{h}_{n,i,j}|^2 - V \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \Big| \hat{\mathbf{H}}\Big] \\ s.t. \quad & \sum_j \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \leq P_{\text{peak}} \qquad \forall i. \end{aligned} \tag{4.66}$$

We evaluate the terms inside the expectation in the following manner.

$$\begin{aligned}\mathbb{E}\big[|\mathbf{w}_{i,j}^H\mathbf{h}_{i,i,j}|^2\big|\hat{\mathbf{H}}\big] &= \mathbb{E}\big[\mathrm{tr}(\mathbf{W}_{i,j}\mathbf{H}_{i,i,j}|)\big|\hat{\mathbf{h}}_{i,i,j}\big] \\ &= \mathrm{tr}(\mathbf{W}_{i,j}\mathbb{E}\big[\mathbf{H}_{i,i,j}\big|\hat{\mathbf{h}}_{i,i,j}\big]) \\ &= \mathrm{tr}(\mathbf{W}_{i,j}(\hat{\mathbf{H}}_{i,i,j}+\sigma_e^2\mathbf{I})).\end{aligned}$$

The last step follows since, for the channel estimation model of (4.61), we have

$$\begin{aligned}\mathbb{E}\left[\mathbf{h}_{i,i,j}\mathbf{h}_{i,i,j}^H\big|\hat{\mathbf{h}}\right] &= \mathbb{E}\left[(\hat{\mathbf{h}}_{i,i,j}+\tilde{\mathbf{h}}_{i,i,j})(\hat{\mathbf{h}}_{i,i,j}+\tilde{\mathbf{h}}_{i,i,j}^H)\big|\hat{\mathbf{h}}_{i,i,j}\right] \\ &= \hat{\mathbf{h}}_{i,i,j}\hat{\mathbf{h}}_{i,i,j}^H + \mathbb{E}\left[\tilde{\mathbf{h}}_{i,i,j}\tilde{\mathbf{h}}_{i,i,j}^H\big|\hat{\mathbf{h}}_{i,i,j}\right] \\ &\overset{(a)}{=} \hat{\mathbf{h}}_{i,i,j}\hat{\mathbf{h}}_{i,i,j}^H + \sigma_e^2\mathbf{I} \qquad (4.67)\end{aligned}$$

where (a) follows since the estimation error is independent of the estimate for MMSE estimation. Similarly, for the interference terms,

$$\mathbb{E}\left[|\mathbf{w}_{n,k}^H\mathbf{h}_{n,i,j}|^2\Big|\hat{\mathbf{H}}\right] = \mathrm{tr}(\mathbf{W}_{n,k}(\hat{\mathbf{H}}_{n,i,j}+\sigma_e^2\mathbf{I})). \qquad (4.68)$$

Using (4.67) and (4.68), the optimization problem (4.66) becomes,

$$\begin{aligned}\max_{\mathbf{w}} \quad & \sum_{i,j}\Big[Q_{i,j}\mathrm{tr}(\mathbf{W}_{i,j}(\hat{\mathbf{H}}_{i,i,j}+\sigma_e^2\mathbf{I})) \qquad (4.69)\\ & -Q_{i,j}\sum_{\substack{(n,k)\\ \neq(i,j)}}\mathrm{tr}(\mathbf{W}_{n,k}(\hat{\mathbf{H}}_{n,i,j}+\sigma_e^2\mathbf{I})) - V\mathbb{E}\mathrm{tr}(\mathbf{W}_{i,j})\Big] \\ s.t. \quad & \sum_j \mathrm{tr}(\mathbf{W}_{i,j}) \le P_{\text{peak}} \qquad \forall i.\end{aligned}$$

Similar to the steps developed in Section 4.4.1, we can rearrange the optimization problem as follows:

$$\begin{aligned}\max_{\mathbf{w},\tau} \quad & \sum_{i,j}\mathrm{tr}(\mathbf{A}_{i,j}^{\text{est}}\mathbf{W}_{i,j}) \qquad (4.70)\\ s.t. \quad & \sum_j \mathrm{tr}(\mathbf{W}_{i,j}) \le P_{\text{peak}} \qquad \forall i\end{aligned}$$

where

$$\mathbf{A}_{i,j}^{\text{est}} = Q_{i,j}(\hat{\mathbf{H}}_{i,i,j}+\sigma_e^2\mathbf{I})) - \sum_{\substack{(n,k)\\ \neq(i,j)}} Q_{n,k}(\hat{\mathbf{H}}_{n,i,j}+\sigma_e^2\mathbf{I}) - V\mathbf{I}. \qquad (4.71)$$

Optimization problem (4.70) can be solved using the method of SDP or eigen decomposition method as in the case with perfect CSI. Similar to the steps followed in the case of perfect CSI, we can also obtain the performance bounds in terms of the average queue-length of the virtaul queue and average energy expenditure for any target QoS lying inside the region defined in (4.65).

4.7 Numerical Results

In this section, we present some numerical results to demonstrate the performance of the DBF algorithm. We consider a system consisting of 2 cells with each cell having 2 UTs each. Each BS has 5 antennas and peak transmit power of 10dB per BS. We consider a distance dependent path loss model in which the UTs are assumed to be arbitrarily scattered inside each cell. The path loss factor from from BS_i to $\text{UT}_{j,k}$ is given as $\sigma_{i,j,k} = d_{i,j,k}^{-\beta}$ where $d_{i,j,k}$ is the distance between BS_i to $\text{UT}_{j,k}$, normalized to the maximum distance within a cell, and β is the path loss exponent which lies usually in the range from 2 to 5 dependent on the radio environment. We also assume that no user terminal is within a normalized distance of 0.1 from the closest BS.

We plot the time average energy expenditure per BS versus the target QoS for two cases. In the first case, we solve the problem of minimizing the instantaneous energy expenditure subject to instantaneous QoS constraints ($\min_{\mathbf{w}} \sum_i P_i[t]$ s.t. $\gamma_{i,j}[t] \geq \lambda_{i,j}\ \forall t$). We repeat this for 1000 time slots. In the second scenario, we solve the problem of minimizing the time average energy expenditure subject to time average QoS constraints ($\bar{\gamma}_{i,j} \geq \lambda_{i,j}$). We plot the result in Figure 4.2. It can be seen that for the case with time average constraints, the energy expenditure is lower. In particular, for a target QoS of 10dB, energy minimization with time average QoS constraints under the Lypunov optimization based approach provides upto 4dB reduction in the energy expenditure as compared to the case with instantaneous QoS constraints (for $V = 800$.) This is in accordance with our intuition that the time average QoS constraint provides greater flexibility in allocating resources over channel fading states.

We next plot the time average energy expenditure per BS versus the average queue-length for different values of V obtained by running the DBF algorithm for 1000 time slots. The target time average QoS is 10dB. It can be seen that as the value of V increases, the time average energy expenditure energy expenditure decreases and the average queue-length increases. This is in accordance with the performance bounds of DBF algorithm. Increasing V implies that the BS transmits less frequently resulting in higher average queue-length and lower average energy expenditure.

Next, we examine the impact of the number of transmit antennas on the target QoS and the energy expenditure. We first plot of average queue-length (of the virtual queue) as a function the target QoS for different number of transmit antennas in Figure 4.4. First, it can be seen that as the number of transmit antennas increase, the average queue-length becomes lower. Also, it

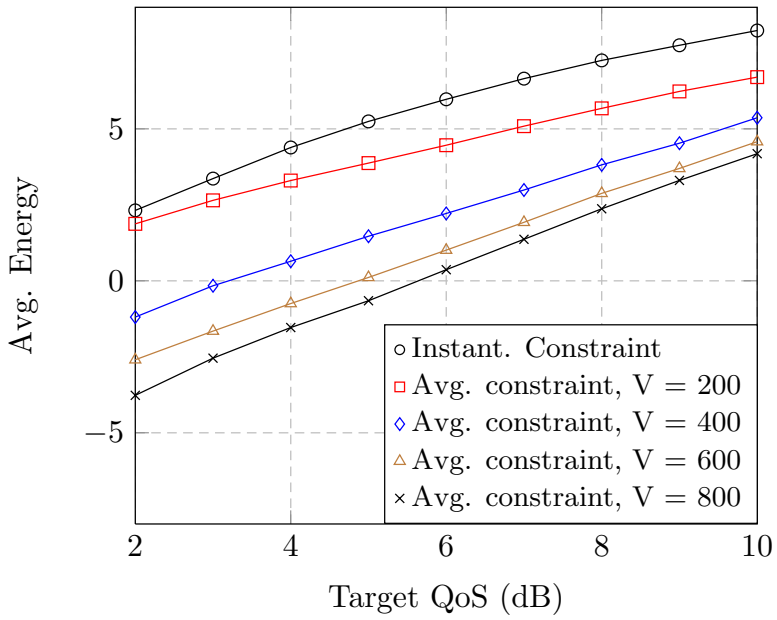

Figure 4.2: Average energy expenditure Vs target QoS for a two cell scenario, each cell consisting of two UTs, $N_t = 5$, $P_{\text{peak}} = 10$dB.

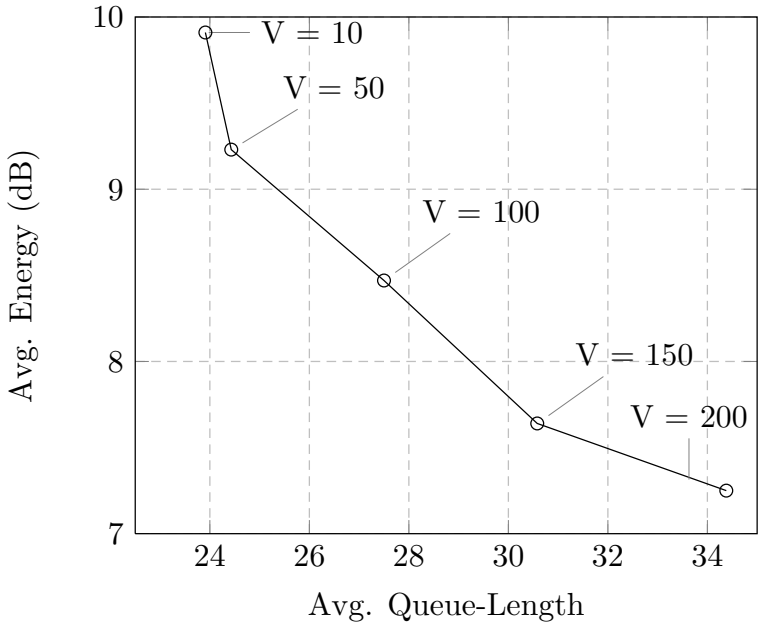

Figure 4.3: Time average queue-length vs time average energy expenditure for a two cell scenario, each cell consisting of two UTs, $N_t = 5$, peak power per BS = 10dB Target QoS value= 10dB.

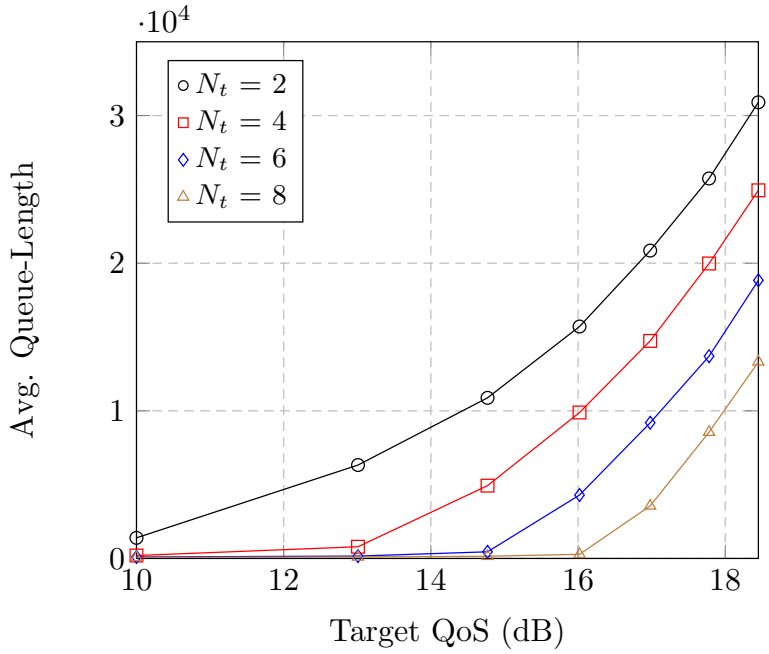

Figure 4.4: Time average queue-length vs target QoS in dB for different number of transmit antennas, peak power per BS = 10dB, V = 100.

can be seen that there is a cut of point beyond which the average queue-length blows up. The cut off point represents the maximum supportable QoS target for the system. Higher the number of transmit antennas, higher is the maximum supportable QoS. This is due to the fact that higher number of antennas leads to greater degrees of freedom resulting in enhancement of the useful signal power and less interference power. We also plot the average energy expenditure for a target QoS of 10dB as a function of number of antennas on the BS in Figure 4.5 . As before, the average energy expenditure reduces with the increase in number of BS antennas.

Finally, we plot the average energy expenditure as a function of the normalized variance of the channel estimation error in Figure 4.6. In this plot, $\sigma_e^2 = 0$ implies perfect knowledge of CSI at the BS and $\sigma_e^2 = 1$ implies that the channel estimate is independent of the actual channel realization. It can be seen that as the quality of the CSI estimate gets worse, the average energy expenditure becomes higher. Therefore, from a purely form energy expenditure point of view, it is necessary to have a good channel estimate. However, it must be noted that our model does not capture the loss in the spectral efficiency of the system incurred while achieving a good channel estimate.

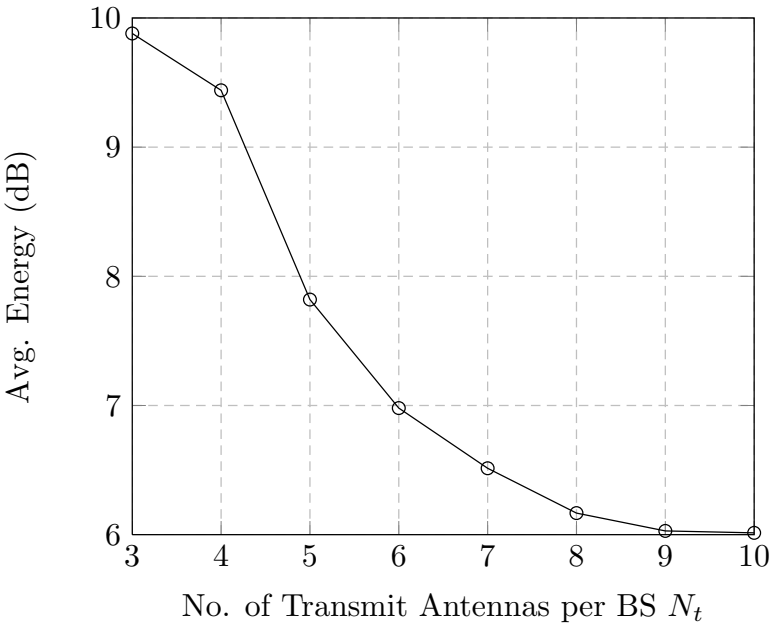

Figure 4.5: Average energy expenditure vs number of transmit antennas, peak power = 10dB, target QoS = 10dB, V = 100.

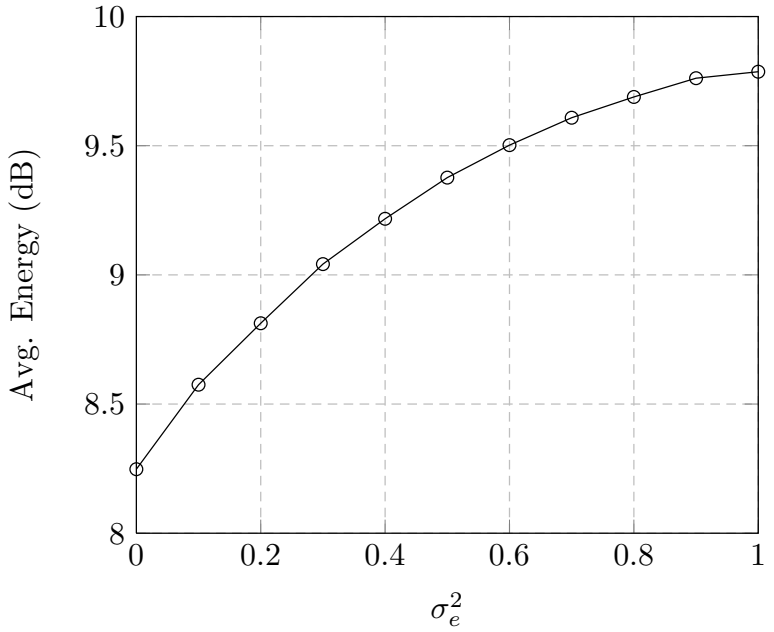

Figure 4.6: Average energy expenditure vs variance of the channel estimation error, peak power = 10dB, target QoS = 10dB, V = 100.

4.8 Conclusion

In this work, we handled the problem of minimizing the time average energy expenditure subject to satisfying time average QoS constraints in a MIMO multi-cell scenario. Using the technique of Lyapunov optimization, we proposed a decentralized online beamforming design algorithm whose performance in terms of the time average expenditure can be made arbitrarily close to the optimal. We also examined the impact of delayed information exchange on the performance of our algorithm. Our results show that time average QoS constraints can lead to better savings in terms of energy expenditure as compared to solving the problem with instantaneous constraints.

4.9 Appendices

4.9.1 Performance Bounds for the DBF Algorithm

Part I: Proof of Proposition 3

From (4.16), we can write the following.

$$\begin{aligned} Q_{i,j}^2[t+1] \leq & \left(Q_{i,j}[t] - \mu_{i,j}[t]\right)^2 + A_{i,j}^2[t] \\ & + 2A_{i,j}[t] \max\left(0, Q_{i,j}[t] - \mu_{i,j}[t]\right) \\ & \leq Q_{i,j}^2[t] + \mu_{i,j}^2[t] + A_{i,j}^2[t] \\ & - 2Q_{i,j}[t]\left(\mu_{i,j}[t] - A_{i,j}[t]\right). \end{aligned} \tag{4.72}$$

Summing with respect to i, j and taking the conditional expectation $\mathbb{E}[.|\mathbf{Q}[t]]$, we have,

$$\Delta(\mathbf{Q}[t]) \leq \sum_{i,j} \mathbb{E}\left[\mu_{i,j}^2[t] + A_{i,j}^2[t] | \mathbf{Q}[t]\right] - \sum_{i,j} Q_{i,j}[t] \mathbb{E}\left[\mu_{i,j}[t] - A_{i,j}[t] | \mathbf{Q}[t]\right] \tag{4.73}$$

Now, we can provide the following bounds,

$$\begin{aligned} A_{i,j}[t] = & \sum_{\substack{(n,k) \\ \neq (i,j)}} |\mathbf{w}_{n,k}^H[t] \mathbf{h}_{n,i,j}[t]|^2 + \lambda_{i,j} \\ & \leq P_{\text{peak}} \sum_n |\mathbf{h}_{n,i,j}[t]|^2 + \lambda_{i,j} \\ & \leq N P_{\text{peak}} \max_n |\mathbf{h}_{n,i,j}[t]|^2 + \lambda_{i,j} \triangleq A_{i,j}^{\max}[t] \end{aligned} \tag{4.74}$$

and

$$\begin{aligned} \mu_{i,j}[t] = & |\mathbf{w}_{i,j}^H[t] \mathbf{h}_{i,i,j}[t]|^2 \\ & \leq P_{\text{peak}} |\mathbf{h}_{i,i,j}[t]|^2 \triangleq \mu_{i,j}^{\max}[t] \end{aligned} \tag{4.75}$$

where the upper bound is derived using $\sum_j \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t] \leq P_{\text{peak}}\ \forall i$ and the fact that for two vectors $\mathbf{x}$ and $\mathbf{y}$, the quantity $|\mathbf{xy}|^2$ is maximized when $\mathbf{x} = \mathbf{y}^H$. Using the bounds of (4.74) and noting that all the quantities are bounded, we have,

$$A_{i,j}^2[t] + \mu_{i,j}^2[t] \leq (A_{i,j}^{\max}[t])^2 + (\mu_{i,j}^{\max}[t])^2. \tag{4.76}$$

From (4.73),(4.76) and substituting the expressions for $A_{i,j}[t]$ and $\mu_{i,j}[t]$, we obtain,

$$\begin{aligned}\Delta(\mathbf{Q}[t]) \leq & \sum_{i,j} \mathbb{E}\Big[(A_{i,j}^{\max}[t])^2 + (\mu_{i,j}^{\max}[t])^2\Big] \\ & + \sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j} \mathbb{E}\Big[Q_{i,j}[t]\Big(|\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2 \\ & - \sum_{\substack{(n,k)\\ \neq(i,j)}} |\mathbf{w}_{n,k}^H[t]\mathbf{h}_{n,i,j}[t]|^2\Big)\Big|\mathbf{Q}(t)\Big]. \end{aligned} \tag{4.77}$$

Adding the term $V\mathbb{E}\big[\sum_{i,j} \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\big|\mathbf{Q}[t]\big]$ to both the sides of the we obtain,

$$\begin{aligned}\Delta(\mathbf{Q}[t]) + V\mathbb{E}\big[\sum_{i,j} & \mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\big|\mathbf{Q}[t]\big] \leq C_1 \\ & + \sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j} \mathbb{E}\Big[Q_{i,j}[t]\Big(|\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2 \\ & \qquad - \sum_{\substack{(n,k)\\ \neq(i,j)}} |\mathbf{w}_{n,k}^H[t]\mathbf{h}_{n,i,j}[t]|^2\Big) \\ & \quad - V\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\Big|\mathbf{Q}[t]\Big]. \end{aligned} \tag{4.78}$$

Part II : Proof of Proposition 5

From (4.78), for the DBF policy we have,

$$\begin{aligned}
&\Delta(\mathbf{Q}[t]) + V\mathbb{E}\big[\sum_{i,j}(\mathbf{w}_{i,j}^{\text{opt}})^H[t]\mathbf{w}_{i,j}^{\text{opt}}[t]\big|\mathbf{Q}[t]\big] \leq C_1 \\
&+\sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j}\mathbb{E}\Big[Q_{i,j}[t]\Big((|\mathbf{w}_{i,j}^{\text{opt}}[t])^H\mathbf{h}_{i,i,j}[t]|^2 \\
&\qquad - \sum_{\substack{(n,k)\\ \neq(i,j)}} |(\mathbf{w}_{n,k}^{\text{opt}}[t])^H\mathbf{h}_{n,i,j}[t]|^2\Big) \\
&\quad - V(\mathbf{w}_{i,j}^{\text{opt}}[t])^H\mathbf{w}_{i,j}^{\text{opt}}[t]\Big|\mathbf{Q}[t]\Big] \qquad (4.79)\\
&\qquad \overset{(a)}{\leq} C_1 \\
&+\sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j}\mathbb{E}\Big[Q_{i,j}[t]\Big(|\mathbf{w}_{i,j}^{\text{TS}}[t])^H\mathbf{h}_{i,i,j}[t]|^2 \\
&\qquad - \sum_{\substack{(n,k)\\ \neq(i,j)}} |(\mathbf{w}_{n,k}^{\text{TS}}[t])^H\mathbf{h}_{n,i,j}[t]|^2\Big) \\
&\quad - V(\mathbf{w}_{i,j}^{\text{TS}}[t])^H\mathbf{w}_{i,j}^{\text{TS}}[t]\Big|\mathbf{Q}[t]\Big] \qquad (4.80)
\end{aligned}$$

where the beamforming vector $\mathbf{w}_{i,j}^{\text{TS}}$ is the one implemented with any stationary randomized policy. Inequality (a) follows due to the following reason. Recall that the DBF algorithm is implemented to maximize the RHS of the bound in (4.79). Therefore, replacing with any other policy should yield the inequality of (a).

In particular we will replace by a stationary randomized policy which satisfies the following conditions.

$$\begin{aligned}
&\mathbb{E}\Big[|(\mathbf{w}^{\text{TS}}{}_{i,j}[t])^H\mathbf{h}_{i,i,j}[t]|^2 - \\
&\quad\sum_{\substack{(n,k)\\ \neq(i,j)}} |(\mathbf{w}^{\text{TS}}{}_{n,k}[t])^H\mathbf{h}_{n,i,j}[t]|^2\big|\mathbf{Q}(t)\Big] \geq \lambda_{i,j} + \epsilon,\ \forall i,j \qquad (4.81)\\
&\mathbb{E}\Big[\sum_{i,j}(\mathbf{w}^{\text{TS}}{}_{i,j}[t])^H\mathbf{w}_{i,j}[t]\big|\mathbf{Q}(t)\Big] = P_{\text{inf}}(\epsilon) \qquad (4.82)
\end{aligned}$$

for some $\epsilon > 0$. The existance of such a policy is proved in [78] for any $\lambda_{i,j}$ lying strictly inside the achievable QoS region. Using (4.81) and (4.82) in (4.80)

yields,

$$\begin{aligned}\Delta(\mathbf{Q}(t)) + V\mathbb{E}\big[\sum_{i,j}(\mathbf{w}_{i,j}^{\text{opt}}[t])^H\mathbf{w}_{i,j}^{\text{opt}}[t]\big|\mathbf{Q}[t]\big] &\leq C_1 \\ + \sum_{i,j} Q_{i,j}(t)\lambda_{i,j} - \sum_{i,j} Q_{i,j}(t)(\lambda_{i,j}+\epsilon) - VP_{\text{inf}}(\epsilon) & \\ = NKB_1 - \epsilon\sum_{i,j} Q_{i,j}(t) - VP_{\text{inf}}(\epsilon) & \qquad (4.83)\end{aligned}$$

From (4.83) and Theorem 7, we can conclude that,

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i,j}\mathbb{E}\left[Q_{i,j}^{\text{opt}}[t]\right] \leq \frac{C_1 + VNKP_{\text{peak}}}{\epsilon}$$

and

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i}\mathbb{E}\left[P_{i}^{\text{opt}}[t]\right] \leq P_{\text{inf}} + \frac{C_1}{V}$$

4.9.2 Performance with Delayed Queue-length Exchange

Part I : Proof of Lemma 12

From the definitions of $\mathbf{W}_{i,j}^{\text{opt}}[t]$ and $\mathbf{W}_{i,j}^{\tau}[t]$, we can conclude the following

$$\sum_j \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{opt}}[t]) \geq \sum_j \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{del}}[t]) \qquad (4.84)$$

$$\sum_j \text{tr}(\mathbf{A}_{i,j}^{\tau}[t]\mathbf{W}_{i,j}^{\text{del}}[t]) \geq \sum_j \text{tr}(\mathbf{A}_{i,j}^{\tau}[t]\mathbf{W}_{i,j}^{\text{opt}}[t]) \qquad (4.85)$$

Recall the expression for $\mathbf{A}_{i,j}[t]$ given by

$$\mathbf{A}_{i,j}[t] = Q_{i,j}[t]\mathbf{H}_{i,i,j}[t] - \sum_{k\neq j} Q_{i,k}[t]\mathbf{H}_{i,i,k}[t] - \sum_{n\neq i,k} Q_{n,k}[t]\mathbf{H}_{i,n,k}[t] - V$$

Adding and subtracting $\sum_{n\neq i,k} Q_{n,k}[t-\tau]\mathbf{H}_{i,n,k}[t]$ on the right hand side, we obtain

$$\begin{aligned}\mathbf{A}_{i,j}[t] = Q_{i,j}[t]\mathbf{H}_{i,i,j}[t] &- \sum_{k\neq j} Q_{i,k}[t]\mathbf{H}_{i,i,k}[t] \\ &- \sum_{n\neq i,k} Q_{n,k}[t-\tau]\mathbf{H}_{i,n,k}[t] \\ &+ \sum_{n\neq i,k}(Q_{n,k}[t-\tau] - Q_{n,k}[t])\mathbf{H}_{i,n,k}[t] - V \\ &= \mathbf{A}_{i,j}^{\tau}[t] + \sum_{n\neq i,k} Q_{n,k}^{d}[t]\mathbf{H}_{i,n,k}[t] \qquad (4.86)\end{aligned}$$

where $Q^d_{n,k}[t] = Q_{n,k}[t-\tau] - Q_{n,k}[t]$. Using (4.86) in the inequality (4.84) yields,

$$\begin{aligned}
&\sum_j \mathrm{tr}(\mathbf{A}^{\tau}_{i,j}[t]\mathbf{W}^{\mathrm{opt}}_{i,j}[t]) \\
&\qquad + \sum_j \sum_{n\neq i,k} Q^d_{n,k}[t]\mathrm{tr}(\mathbf{H}_{i,n,k}[t]\mathbf{W}^{\mathrm{opt}}_{i,j}[t]) \\
&\qquad \geq \sum_j \mathrm{tr}(A^{\tau}_{i,j}[t]\mathbf{W}^{\mathrm{del}}_{i,j}[t]) \\
&\qquad + \sum_j \sum_{n\neq i,k} Q^d_{n,k}[t]\mathrm{tr}(\mathbf{H}_{i,n,k}[t]\mathbf{W}^{\mathrm{del}}_{i,j}[t]) \qquad (4.87)
\end{aligned}$$

Adding (4.85) to (4.87) yields,

$$\sum_j \sum_{n\neq i,k} Q^d_{n,k}[t]\mathrm{tr}(\mathbf{H}_{i,n,k}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \geq 0. \qquad (4.88)$$

Finally consider the term $\sum_j \mathrm{tr}(\mathbf{A}_{i,j}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t]))$. Using (4.86), we obtain

$$\begin{aligned}
&\sum_j \mathrm{tr}(\mathbf{A}_{i,j}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \\
&= \sum_j \mathrm{tr}(\mathbf{A}^{\tau}_{i,j}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \\
&\quad + \sum_j \sum_{n\neq i,k} Q^d_{n,k}[t]\mathrm{tr}(\mathbf{H}_{i,n,k}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \\
&\quad \leq \sum_j \sum_{n\neq i,k} Q^d_{n,k}[t]\mathrm{tr}(\mathbf{H}_{i,n,k}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \qquad (4.89)
\end{aligned}$$

where the last inequality follows due to the fact that (from (4.84))

$$\sum_j \mathrm{tr}(\mathbf{A}^{\tau}_{i,j}[t](\mathbf{W}^{\mathrm{opt}}_{i,j}[t] - \mathbf{W}^{\mathrm{del}}_{i,j}[t])) \leq 0.$$

Finally, we will show that the term on the right hand side of (4.89) is bounded by a constant which is independent of the current queue-lengths. We proceed as follows. Using the equation for queue-length evolution in (4.16) and the bounds in (4.74), we can conclude the following.

$$Q_{i,j}[t-\tau] + \tau A_{\max} \geq Q_{i,j}[t] \qquad \forall i,j \qquad (4.90)$$

$$Q_{i,j}[t-\tau] - \tau \mu_{\max} \geq Q_{i,j}[t] \qquad \forall i,j \qquad (4.91)$$

Combining (4.90) and (4.91) yields,

$$Q_{i,j}[t] - \tau A_{\max} \leq Q_{i,j}[t-\tau] \leq Q_{i,j}[t] + \tau \mu_{\max}, \quad \forall i,j \qquad (4.92)$$

And hence we can conclude that

$$-\tau A_{\max} \leq Q^d_{i,j}[t] \leq \tau \mu_{\max} \qquad \forall i,j \qquad (4.93)$$

Note that in (4.88), we have already shown that the right hand side of (4.89) is positive.

In order to bound the right hand side of (4.89), we proceed as follows. First, we recall that there can only be one active UT per cell. Therefore, only one of $\mathbf{W}_{i,j}^{\text{opt}}$ is a non-zero matrix. Similarly, the case for $\mathbf{W}_{i,j}^{\text{del}}$. Therefore,

$$\begin{aligned}&\sum_j \sum_{n\neq i,k} Q_{n,k}^d[t]\text{tr}(\mathbf{H}_{i,n,k}[t](\mathbf{W}_{i,j}^{\text{opt}}[t]-\mathbf{W}_{i,j}^{\text{del}}[t]))\\ &\quad\leq 2\max\{\tau A_{\max},\tau\mu_{\max}\}P_{\text{peak}}\max_{n\neq i,k}|h_{i,n,k}[t]|^2 \triangleq B_1 \qquad (4.94)\end{aligned}$$

Therefore,

$$\sum_j \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{opt}}[t]) \leq \sum_j \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{del}}[t]) + B_1 \qquad \forall i,t. \qquad (4.95)$$

and hence,

$$\sum_{i,j} \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{opt}}[t]) \leq \sum_j \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{del}}[t]) + C_2 \qquad \forall t. \qquad (4.96)$$

where $C_2 = NB_1 < \infty$.

Part II : Proof of Theorem 8

Rewriting (4.78), we have

$$\begin{aligned}&\Delta(\mathbf{Q}[t]) + V\mathbb{E}\big[\sum_{i,j}\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\big|\mathbf{Q}[t]\big] \leq C_1\\ &\quad+\sum_{i,j}Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j}\mathbb{E}\Big[Q_{i,j}[t]\Big(|\mathbf{w}_{i,j}^H[t]\mathbf{h}_{i,i,j}[t]|^2\\ &\qquad\qquad-\sum_{\substack{(n,k)\\\neq(i,j)}}|\mathbf{w}_{n,k}^H[t]\mathbf{h}_{n,i,j}[t]|^2\Big)\\ &\qquad-V\mathbf{w}_{i,j}^H[t]\mathbf{w}_{i,j}[t]\Big|\mathbf{Q}[t]\Big]\\ &\quad= C_1 + \sum_{i,j}Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j}\text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}[t]) \qquad (4.97)\end{aligned}$$

where the last step follows from the equivalence of the quadratic form and the trace form (as shown in (4.40)). With delayed queue-length information, the

policy corresponding to $\mathbf{W}^{\text{del}}$ is implemented. Therefore, we have

$$\begin{aligned}
&\Delta(\mathbf{Q}(t)) + V\mathbb{E}\big[\sum_{i,j}(\mathbf{w}_{i,j}^{\text{del}}[t])^H\mathbf{w}_{i,j}^{\text{del}}[t]\big|\mathbf{Q}[t]\big] \\
&\leq C_1 + \sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \sum_{i,j}\text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{del}}[t]) \\
&\overset{(a)}{\leq} C_1 + \sum_{i,j} Q_{i,j}[t]\lambda_{i,j} \\
&\qquad\qquad - (\text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{opt}}[t]) - C_2) \\
&= C_1 + C_2 + \sum_{i,j} Q_{i,j}[t]\lambda_{i,j} - \text{tr}(\mathbf{A}_{i,j}[t]\mathbf{W}_{i,j}^{\text{opt}}[t])
\end{aligned} \tag{4.98}$$

where (a) follows from Lemma 12. Now following similar steps as Subsection 4.9.1, part II, we can obtain the bounds of (4.59) and (4.60).

Chapter 5

Decentralized Scheduling Policies

Abstract: *There has been substantial interest over the last decade in developing low complexity decentralized scheduling algorithms in wireless networks. In this context, the queue-length based CSMA scheduling algorithms have attracted significant attention because of their attractive throughput guarantees. However, the CSMA results rely on the mixing of the underlying Markov chain and their performance under fading channel states is unknown.*

In this work, we formulate a partially decentralized randomized scheduling algorithm for a two transmitter receiver pair set-up and investigate its stability properties. Our work is based on the Fast-CSMA (FCSMA) algorithm first developed in [86]. In this work, we generalize the FCSMA based scheduling algorithm to an SINR based interference model in which one or more transmitters can transmit simultaneously while causing interference to the other. We first show that straightforward application of the FCSMA algorithm to an SINR based interference model has low performance. In order to improve the performance of the system, we split the traffic arriving at the transmitter into schedule based queues *and combine it with the FCSMA based scheduling algorithm. We theoretically examine the performance of our algorithm in both non-fading and fading environments and characterize the set of arrival rates which can be stabilized by our proposed algorithm.*

5.1 Introduction

We consider the problem of decentralized channel access for the two user interference channel. Classical information theoretic approach assumes that the transmitters are always saturated with information bits. However, in this work, we consider the randomness in arrival of information bits and hence account

for the queuing backlog at the transmitters. We consider that the transmission rates of each transmitter-receiver pair are a function of the SINR at the receiver.

The work here is comparable to the stream of works related to scheduling algorithms in wireless networks which assume that a fixed number of packets can be transmitted per time slot. The problem then is to schedule a set of non-conflicting links for transmission (*conflict graph* based interference model) in order to ensure stability[1] of the associated queues in the network. The authors in the seminal works [11, 87] developed a maximum-weight based scheduling strategy which is proved to be throughput-optimal. The max-weight result was generalized by many works that followed [12, 88, 89]. However, the max-weight based algorithms are centralized in nature and suffer from high computational complexity. Subsequently low-complexity, decentralized, and possibly suboptimal scheduling algorithms were developed in series of works [90, 91, 92, 93, 94, 95] with varying complexities and performances.

Recently, a class of randomized scheduling algorithms namely the CSMA-based scheduling algorithms have received a lot of attention because of their attractive throughput guarantees. In particular, the authors in [96, 97] proposed an algorithm that adaptively selects the CSMA parameters as a function of the current queue-length and showed that such an algorithm is throughput-optimal. However, this algorithm was based on the time-scale separation assumption, i.e., the Markov Chain underlying the CSMA-based algorithm quickly converges to steady-state as compared to the time-scale of updating parameters of the algorithm. The CSMA based scheduling algorithm was generalized in subsequent works [98, 99]. However, all of the CSMA based scheduling algorithms rely on the mixing of the underlying Markov chain which cannot be guaranteed in a fading environment. Hence, their performance in fading environment is not known.

In this work, we develop a FCSMA (Fast-CSMA) based scheduling algorithm for the SINR-based interference model. The FCSMA operation has advantage over the CSMA based scheduling algorithms under fading channel conditions in that it quickly reaches one of the favorable schedules and sticks to it rather than relying on the convergence of the underlying Markov chain. Hence, the FCSMA based algorithm can perform well under fading environment as well.

We first note that the straightforward application of FCSMA to the SINR based interference model has a low performance. In order to improve the performance of this scheme, we formulate a dynamic rule to split the incoming traffic into *schedule based queues* at the transmitters and combine it with the FCSMA

[1]The exact notion of stability used will be specified later in the chapter.

scheduling. By favorably tuning the control parameter of the traffic splitting rule, we prove that the FCSMA based algorithm along with the appropriate traffic splitting rule can provide a good performance.

Finally, we mention reference [100] for a decentralized queue-length dependent probabilistic scheduling algorithm for the two user multi-access channel. However, the analysis of the algorithm is done assuming that the channel realization stays constant and hence assumes a non-fading scenario. In contrast, we analyze our system under fading environment as well.

5.2 System Model

We consider a set up in which two transmitters (Tx) are trying to communicate to their respective receivers (Rx) over a common resource block as shown in Figure 5.1.

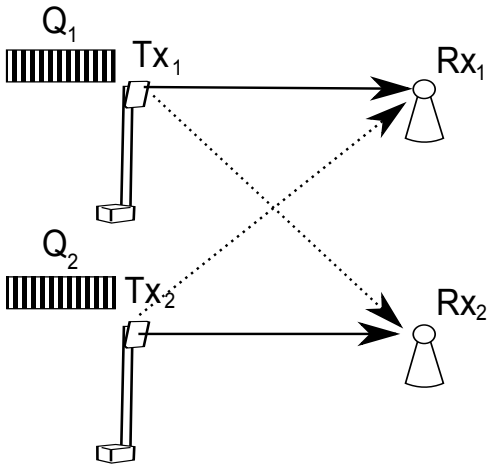

Figure 5.1: A two transmitter-receiver scenario

We assume that the system operates in a time slotted fashion. We denote $A_i[t]$ as the amount of information bits that flows into the Tx_i during each time slot t. The arrival process is assumed to be independent across the users and identically distributed over time slots with a rate of $\lambda_i,\ i = 1, 2$, and $A_i[t] \leq K,\ \forall t$, for some $0 < K < \infty$. Accordingly, there is a queue associated with Tx_i whose queue-length at time slot t is denoted by the notation $Q_i[t]$. Let $S_i[t]$ denote the number of information bits served from the queue of Tx_i during the time slot t. The equation for the queue-length evolution is given by

$$Q_i[t+1] = Q_i[t] + A_i[t] - S_i[t] + U_i[t], \tag{5.1}$$

where $U_i[t]$ denotes the unused service, $0 < U_i[t] \leq 1$ if user i is selected for service and $Q_i[t] \leq 1$, $U_i[t] = 0$ otherwise. We say that a queue is stable if

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[Q_i[t]\right] < \infty. \tag{5.2}$$

We consider the SINR based interference model in which one or more transmitters can transmit simultaneously. In this case, the maximum achievable transmission rate for any Tx-Rx depends on the SINR at the Rx. In general, the transmission rate for a Tx-Rx pair during any time slot can be chosen from a continuous set. However, in order to simplify the analysis, we allow two levels of rates for every Tx-Rx pair. First, a rate of R_i when only one of the two transmitters is transmitting (while the other transmitter is turned off) and a rate r_i when both the transmitters transmitting simultaneously (in which case, they cause interference to each other). These rates correspond to the three possible scheduling decisions in the set $\mathbf{\Omega} = \{\omega_1, \omega_2, \omega_3\}$ where the rates obtained in the three scheduling decisions are given by $\{R_1, 0\}, \{0, R_2\}, \{r_1, r_2\}$ respectively. A reasonable assumption is that the maximum achievable rate is an increasing function of the SINR[2]. Hence, we assume that the rates $r_1 \leq R_1$ and $r_2 \leq R_2$. The stability region for this system can be given as the convex hull of the possible transmission rates.

$$\begin{aligned}\mathbf{\Lambda} = \Big\{ & \lambda_1 < \pi_1 R_1 + \pi_3 r_1, \quad \lambda_2 < \pi_2 R_2 + \pi_3 r_2 \\ & \sum_{i=1}^{3} \pi_i = 1, \ \pi_i \geq 0 \Big\}.\end{aligned} \tag{5.3}$$

The stability region of the system is shown in Figure 5.2. Additionally, we note the condition

$$\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1, \tag{5.4}$$

which ensures that the stability region goes beyond the time sharing region.

Comment on the System Model: The system model considered in this work is quite general in terms of the transmission strategy adopted. The multiple rate levels could be correspond to strategies such as different powers levels of operation, different beamforming strategies adopted at the transmitter, different modulation and coding schemes etc. Our model is quite general to encompass any of these strategies. However, to make the system amenable for analysis, we restrict ourselves to only two rate levels for each transmission.

[2]Note that the rate function is general and not restricted to Shannon capacity rate - $log(1 + SINR)$.

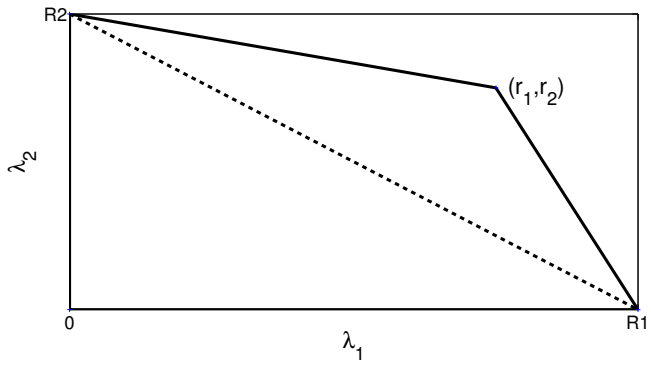

Figure 5.2: Stability region for the 2 User System

5.3 Existing Approaches

In this section, we briefly describe two approaches existing in the literature which have throughput optimality characteristics for the scenario considered in the previous section, namely the maximum weight based scheduling and CSMA based scheduling algorithm respectively.

5.3.1 Maximum Weight based Scheduling Algorithm

A centralized policy which requires the transmitters to share the full CSI is the classic max-weight based policy [11, 12]. Under this policy the scheduler chooses a feasible schedule according to the following rule given as,

$$\boldsymbol{\omega}[t] = \arg\max\{Q_1[t]R_1, Q_2[t]R_2, Q_1[t]r_1 + Q_2[t]r_2\}. \tag{5.5}$$

The throughput-optimality characteristics of max-weight scheduling in both fading and non-fading environments is a well know result [12]. However, max-weight scheduling requires a central entity which has the knowledge of all the channel state and the queue-length information.

5.3.2 CSMA based Scheduling Algorithm

A decentralized scheduling algorithm with throughput optimality characteristics is the CSMA based scheduling algorithm under SINR based interference model [101, 102]. However, we would like to mention at this stage that this algorithm is not applicable under fading channel conditions.

The algorithm works as follows. Each transmitter maintains two backoff timers which are exponential random variables with mean $1/R_i$ and $1/r_i$ $i =$

$1, 2$ (corresponds to the two transmission rates at which they can transmit) respectively. When any one of the timer expires, the corresponding transmitter starts transmitting at the appropriate rate as long as the following conditions are satisfied.

- The transmission itself can tolerate the interference from other active links.
- The transmission would not cause the on-going transmissions of other links to fail.

In other words, a link cannot begin its transmission if its interference cannot be tolerated by any nearby active link, which can in turn ensure that the second requirement be met. For example, if Tx_1 is transmitting at a rate R_1, it implies that the it cannot tolerate any interference from Tx_2 (since the only feasible schedule is $\{R_1, 0\}$) and hence Tx_2 cannot start its transmission (at any of two rates). However, if Tx_1 is transmitting at a rate r_1, Tx_2 is allowed to start simultaneous transmission at a rate r_2 (since the rate $\{r_1, r_2\}$ is feasible). In order to accomplish this, the authors assume that each transmitter always knows how much interference the other active transmitter can tolerate currently. The authors propose that each active Tx updates its interference tolerance level and broadcasts the updated interference tolerance level to neighboring Tx, whenever it starts its transmission. The data transmission time is exponentially distributed with mean 1.

The network state dynamics can be captured by a continuous time Markov chain (CTMC), with each state in the Markov chain corresponding to a feasible state. The operation of the Markov chain satisfies the following conditions.

- State transitions take place from one feasible state to another feasible state where there is only one link state in difference between the two state vectors.
- The duration at each feasible state is exponentially distributed.

The resulting CTMC in the steady state is shown in the Figure 5.3.2. Let us denote $\pi_{\{i,j\}}$ as the steady state probability of being in state $\{i, j\}$. When the

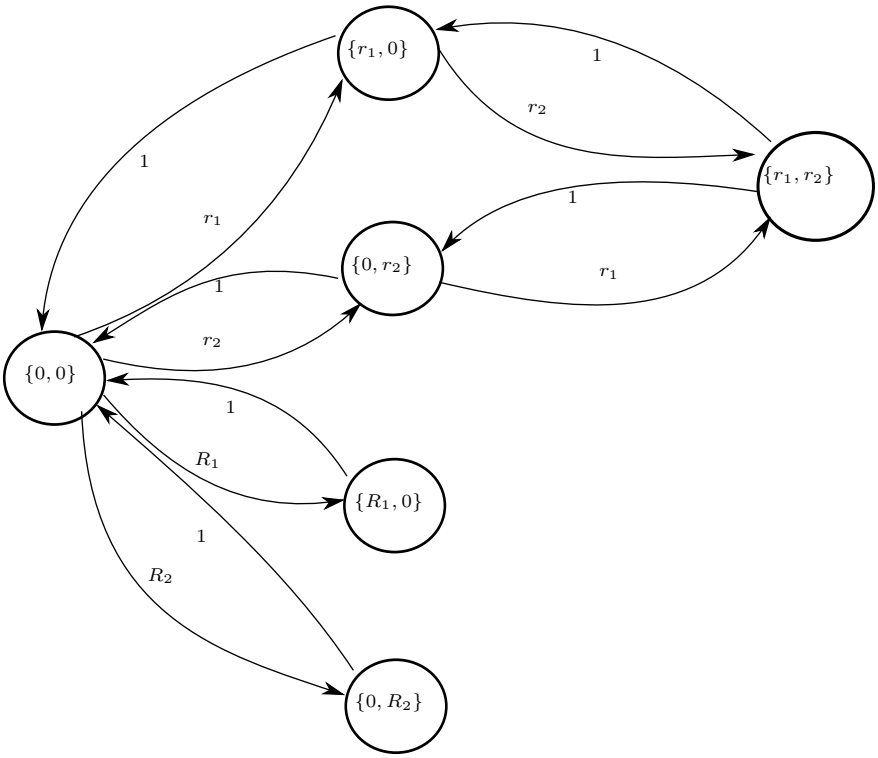

Figure 5.3: CMTC modeling of the scheduling decisions. The scripts on the arrows indicate the transition rates.

Markov chain reaches a stationary distribution, the flow balance equations are

$$\begin{aligned}
\pi_{\{0,0\}} R_1 &= \pi_{\{R_1,0\}} \\
\pi_{\{0,0\}} R_2 &= \pi_{\{0,R_2\}} \\
\pi_{\{0,0\}} r_1 &= \pi_{\{r_1,0\}} \\
\pi_{\{0,0\}} r_2 &= \pi_{\{r_2,0\}} \\
\pi_{\{r_1,0\}} r_2 &= \pi_{\{r_1,r_2\}} \\
\pi_{\{0,r_2\}} r_1 &= \pi_{\{r_1,r_2\}} \\
\pi_{\{0,0\}} + \pi_{\{R_1,0\}} + \pi_{\{0,R_2\}} + \pi_{\{r_1,0\}} + \pi_{\{0,r_2\}} + \pi_{\{r_1,r_2\}} &= 1.
\end{aligned}$$

Solving the above equations, the steady state probabilities are given by

$$\begin{aligned}
\pi_{\{0,0\}} &= \frac{1}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2} \\
\pi_{\{r_1,0\}} &= \frac{r_1}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2} \\
\pi_{\{r_2,0\}} &= \frac{r_2}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2} \\
\pi_{\{R_1,0\}} &= \frac{R_1}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2} \\
\pi_{\{R_2,0\}} &= \frac{R_2}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2} \\
\pi_{\{r_1,r_2\}} &= \frac{r_1 r_2}{1 + R_1 + R_2 + r_1 + r_2 + r_1 r_2}.
\end{aligned}$$

With this background, the online throughput-optimal scheduling algorithm can be developed along the same lines as in [97].

However, as mentioned before, the CSMA based scheduling algorithm relies critically on the mixing of the underlying Markov chain which cannot be guaranteed in a fading environment. Hence its performance in fading environment is not known. In order to overcome the above requirement, in the subsequent part of the chapter, we propose the FCSMA algorithm for SINR based interference model and theoretically examine its performance.

5.4 FCSMA Algorithm Description

We will first describe the FCSMA algorithm for a non fading channel scenario and provide analysis for the same. Later on in Section 5.6, we will take up the FCSMA operation under fading channel scenario. The FCSMA based scheduling algorithm operates in the following way. It operates in a slotted time fashion. At the beginning of time slot t, each Tx independently generates two timers whose values are exponentially distributed random variables with mean $Q_i[t]R_i$

and $Q_i[t]r_i$ respectively. These timers correspond to the respective scheduling decisions in which the Tx_i can achieve a non zero rate. We assume that each Tx maintains a one bit index for the timers associated with it. Let us assume that the index of 0 corresponds to the timer $Q_i[t]R_i$ and an index of 1 indicates that timer $Q_i[t]r_i$.

The system has four timers. Without the loss of generality, assume that one of the timers associated with Tx_1 expires first among the four timers. The algorithm operates in the following manner. Tx_1 immediately suspends its second timer (which has not yet expired) and starts to transmit bits from its queue at the appropriate rate (rate R_1 if timer 0 expires or a rate r_1 if timer 1 expires). Tx_1 communicates the index of timer which has expired to Tx_2. Upon receiving the index bit, Tx_2 also suspends both its timers. We assume the following pre-agreed protocol between the two Txs. Upon reception of the index 0, the Tx_2 keeps silent during corresponding time slot t. Upon reception of the index 1, the Tx_2 transmits from its queue at the rate r_2. We ignore the overhead associated with communicating the bit between the two Txs. The state diagram for the FCSMA based scheduling algorithm is shown in Figure 5.4 and the probabilities of reaching each of the three possible schedules during a time slot t are given by

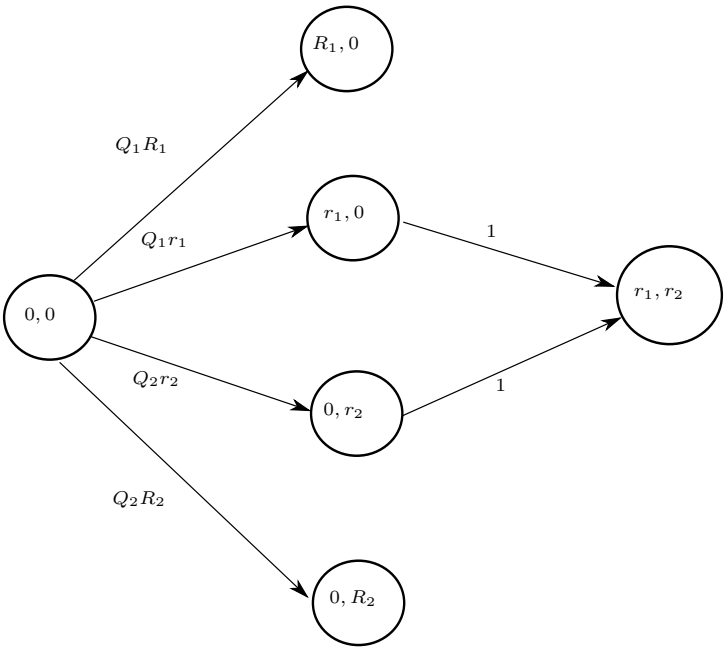

Figure 5.4: FCSMA State Diagram

$$\mathbb{P}_{\omega_1}(\mathbf{Q}) = \frac{R_1 Q_1}{\sum_{k=1}^{2} Q_k(R_k + r_k)},$$
$$\mathbb{P}_{\omega_2}(\mathbf{Q}) = \frac{R_2 Q_2}{\sum_{k=1}^{2} Q_k(R_k + r_k)}$$
$$\mathbb{P}_{\omega_3}(\mathbf{Q}) = \frac{\sum_{k=1}^{2} r_k Q_k}{\sum_{k=1}^{2} Q_k(R_k + r_k)}.$$

Additionally, the expected value of service rate for the queue at Tx_i during the time slot t can be given by

$$\mathbb{E}\left[S_i[t] \middle| \mathbf{Q}[t] = \mathbf{Q}\right] = \frac{R_i^2 Q_i + r_i \sum_{k=1}^{2} r_k Q_k}{\sum_{k=1}^{2} Q_k(R_k + r_k)},\ i = 1, 2. \tag{5.6}$$

Differences with the CSMA Algorithm

At this stage, we point out the difference between the operation of CSMA based scheduling algorithm described in Subsection 5.3.2, and the FCSMA algorithm. The FCSMA algorithm operates in a slot by slot fashion rather than the continuous time fashion. During each time slot, the scheduling algorithm chooses one feasible schedule (which is determined by the first timer that expires during the time slot[3] as described before) and sticks to the schedule during the entire time slot. Hence, the state diagram modeling in Figure 5.4 does not permit backward transitions. At the beginning of the new time slot, the transmitters suspend the previous schedules and choose a new schedule according to the FCSMA operation. Therefore, unlike the CSMA based scheduling algorithm, the FCSMA operation does not rely on the convergence of the underlying Markov chain to its steady state distribution. Rather, during every time slot, it quickly chooses a schedule and sticks to it during the entire time slot. Owing to the quick operation of the FCSMA algorithm, it can be implemented in the fading channel scenario by repeating the FCSMA algorithm during every coherence interval.

The throughput optimality characteristics of FCSMA algorithm proved in [86] for a conflict graph based interference model in which only one transmitter can transmit at a time. In what follows, we will examining the performance of the FCSMA algorithm for an SINR based interference model.

Proposition 6. *Consider a 2-user perfectly symmetric network in which* $R_1 = R_2 = 1$ *and* $r_1 = r_2 = \alpha$. *(Note* $0 \leq \alpha \leq 1$.*) When the mean rate of the arrival process into the two Txs are the same (i.e.,* $\lambda_1 = \lambda_2 = \lambda$*), the maximum arrival*

[3]We assume that the time taken for the first timer to expire is very small compared to the coherence time of the channel.

rate which can be supported by the FCSMA scheduling algorithm is given by

$$\lambda < \frac{\alpha^2}{\alpha+1} + \frac{1}{2(\alpha+1)}. \tag{5.7}$$

The proof of the above proposition can be found in Section 5.8.1.

Remarks on the FCSMA algorithm: The FCSMA based scheduling algorithm described above is a partially decentralized algorithm in which the Txs exchange one bit information (index of the timer that expires first). This calls for a substantially less overhead of information exchange between the Txs as compared to exchanging the full CSI. From the plot of the stability region in Figure 5.2, notice that the maximum achievable rate in the symmetric case $\lambda_1 = \lambda_2 = \lambda$ is $\lambda < \alpha$ (any arrival rate arbitrarily close to λ can be stabilized). However, the bound specified in (5.7) is lesser than α. In what follows, we overcome this problem by combining the FCSMA scheduling scheme with a dynamic traffic splitting algorithm. The main idea behind splitting the incoming traffic is to modify the scheduling probabilities of the FCMSA algorithm in a such a way that the service rates adapt to the incoming traffic arrival rate appropriately.

5.5 FCSMA with Dynamic Traffic Splitting Algorithm

In this section, we introduce the concept of *schedule based queues* to split the input traffic arriving into the Txs. The concept is illustrated in Figure 5.5. Each Tx maintains two different queues one for each scheduling decision. For the Tx_i, $(i = 1, 2)$ the queue Q_{ii} corresponds to the first scheduling decision in which the Tx_i can transmit at the higher rate R_i. When selected for service, this queue gets a service rate of R_i. Let us define $\bar{i} = \mod(i, 2) + 1$ (index of the other user). The second queue $Q_{i\bar{i}}$ corresponds to the scheduling decision in which both the Txs have joint access to the channel and when selected for service, gets a rate of r_i. The traffic splitting policy can be described as follows. During the time slot t, each transmitter compares the current queue-lengths $Q_{ii}[t]$ and $\delta_i Q_{i\bar{i}}[t]$ where $\delta_i \geq 0$ is a scaling factor. If $Q_{ii}[t] < \delta_i Q_{i\bar{i}}[t]$, the information bits arriving in the respective slot enter the queue Q_{ii} and vice versa. Accordingly,

$$\begin{aligned}
\lambda_{ii} = \mathbb{E}[A_{ii}[t]] &= \begin{cases} \lambda_i & \text{if } \delta_i Q_{i\bar{i}}[t] > Q_{ii}[t] \\ 0 & \text{else} \end{cases} \\
\lambda_{i\bar{i}} = \mathbb{E}[A_{i\bar{i}}[t]] &= \begin{cases} \lambda_i & \text{if } \delta_i Q_{i\bar{i}}[t] \leq Q_{ii}[t] \\ 0 & \text{else.} \end{cases}
\end{aligned} \tag{5.8}$$

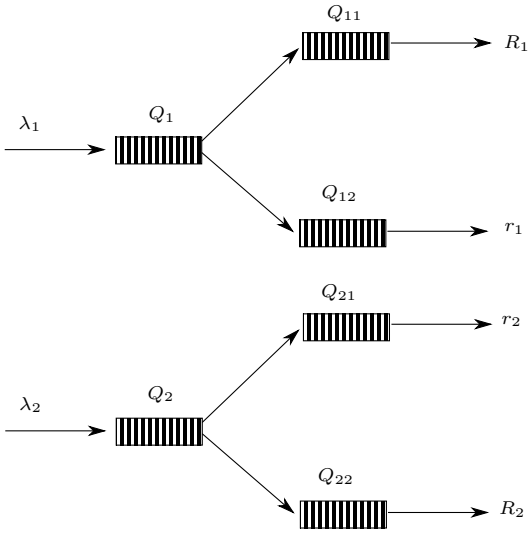

Figure 5.5: Concept of schedule based queues

The scheduling algorithm is exactly the same as the FCSMA algorithm described in Section II B except that the two timers associated with the Tx_i are exponential random variables with mean $Q_{ii}[t]R_i$ and $Q_{i\bar{i}}[t]r_i$ respectively (note that the queue-length values associated with the two mean values are different). The probabilities of each scheduling decision in this case are

$$\mathbb{P}_{\omega_1}(\mathbf{Q}) = \frac{Q_{11}R_{11}}{\sum_{k=1}^{2} Q_{kk}R_k + Q_{k\bar{k}}r_k} \tag{5.9}$$

$$\mathbb{P}_{\omega_2}(\mathbf{Q}) = \frac{Q_{22}[t]R_{22}}{\sum_{k=1}^{2} Q_{kk}R_k + Q_{k\bar{k}}r_k}$$

$$\mathbb{P}_{\omega_3}(\mathbf{Q}) = \frac{\sum_{k=1}^{2} Q_{k\bar{k}}r_k}{\sum_{k=1}^{2} Q_{kk}R_k + Q_{k\bar{k}}r_k}. \tag{5.10}$$

Also, the expected service rate for each queue is given by

$$\mathbb{E}\left[S_{ii}[t]\middle|\mathbf{Q}[t]=\mathbf{Q}\right] = \frac{Q_{ii}R_i^2}{\sum_{k=1}^{2}\left(Q_{kk}R_i + Q_{k\bar{k}}r_k\right)}$$

$$\mathbb{E}\left[S_{i\bar{i}}[t]\middle|\mathbf{Q}[t]=\mathbf{Q}\right] = \frac{r_i\sum_{k=1}^{2} Q_{k\bar{k}}r_k}{\sum_{k=1}^{2}\left(Q_{kk}R_k + Q_{k\bar{k}}r_k\right)}, \qquad i=1,2.$$

Having defined a dynamic traffic splitting policy described above, the next task is to examine theoretically the set of arrival rate which can be can be stabilized by our algorithm. To do the same, we define a Lyapunov function and examine its properties for different values of the queue-lengths. In order to make things more amenable for theoretical analysis, we restrict our proofs to a perfectly symmetric system model.

Theorem 9. *Consider a 2-user perfectly symmetric network described in Proposition 6. When the mean rate of the arrival process into the two transmitters are the same (i.e., $\lambda_1 = \lambda_2 = \lambda$), the maximum arrival rate which can be supported by the traffic splitting policy described in equation* (5.8) *followed by the FCSMA scheduling algorithm is given by*

$$\lambda < \frac{\alpha^2}{\alpha + \delta} + \frac{\delta}{2(\alpha + \delta)}. \tag{5.11}$$

The proof of this theorem in provided in Section 5.8.2.

Remarks on FCSMA with the dynamic traffic splitting algorithm: Note that the result of Theorem 9 can be generalized to the case when the data rates are $\{R, 0\}, \{r, r\}, \{0, R\}$. In this case any rate $\lambda < \frac{r^2}{r+\delta R} + \frac{\delta R^2}{2(r+\delta R)}$ can be stabilized. The FCSMA based algorithm along with the dynamic traffic splitting algorithm provides us with a tunable parameter δ which can be varied in order to achieve better performance. Specifically when $\lambda_1 = \lambda_2 = \lambda$, by setting $\delta = 0$, from the result of Theorem 9 that any rate $\lambda < \alpha$ can be stabilized by the system (which is also the maximum achievable rate in the symmetric arrival case from the plot of the stability region). The parameter δ can be calculated based on the point inside the stability region in which we are operating. Our theoretical analysis was limited to the case of symmetric arrivals ($\lambda_1 = \lambda_2 = \lambda$).

5.6 Fading Channels

Now consider a symmetric block fading model where the channel realization is fixed for a given duration of time (coherence time) but changes independently at every time slot t. The set of channels in the network can assume a state $s = \{1, \ldots, S\}$ according to stationary probability p_s. We denote the cardinality of the set by $|S|$ and $\sum_{s=1}^{|S|} p_s = 1$. In each time slot t, the achievable rate for the three possible scheduling decisions are $\{\{R_s, 0\}, \{r_s, r_s\}, \{0, R_s\}\}$ if the network is in fading state s at time slot t.

The FCSMA algorithm is naturally applicable under fading channel scenario due to the fact that it does not have a Markov chain based operation (and hence does not have to rely on the convergence). Instead, once the first timer expires (which as we mentioned before, assume that the time taken for the first timer to expire is very small compared to the coherence time of the channel), the FCSMA algorithm sticks to the schedule during the entire time slot. Therefore, the FCSMA operation can be repeated during every coherence interval. The following is the mean rate stabilize by FCSMA policy under fading channel scenario.

Theorem 10. *In this scenario, when $\lambda_1 = \lambda_2 = \lambda$, the maximum rate that can be stabilized by the FCSMA policy along with traffic splitting algorithm is given by*

$$\lambda < \sum_{s=1}^{|S|} p_s \left(\frac{r_s^2}{r_s + \delta R_s} + \frac{\delta R_s^2}{2(r_s + \delta R_s)} \right). \tag{5.12}$$

The proof is provided in Section 5.8.3.

5.7 Conclusion

In this work, we have formulated a partially decentralized randomized scheduling algorithm for a two Tx-Rx set up under an SINR based interference model. In our algorithm, the transmitters have to exchange only one bit information between themselves. Our algorithm has advantage over existing scheduling algorithms since it is decentralized in nature and can perform well under fading conditions. Our theoretical analysis was limited to the case of symmetric network settings and symmetric arrival rates.

5.7.1 Future Directions

Comments on generalizing the FCSMA Algorithm

In this work, we have restricted the theoritical analysis of the FCSMA algorithm to a symmetric system set up. The concept of FCSMA can be extended to a more general asymmetric system setup and also to include advanced physical layer techniques such as modulation schemes, power control or beamforming etc. The aforementioned cases would give rise to multiple rate points on the boundary of the stability region as shown in Figure 5.6. Each point $\{r_{1i}, r_{2j}\}$ on the boundary of the stability region corresponds to a rate achievable under some feasible policy (power allocation, beamfomring etc). One can think of extending the FCSMA algorithm in a similar manner by using timers (whose parameters are a function of the current queue-length) to determine a feasible schedule during each time slot. However, the optimal traffic splitting policy and charecterization of the part of the stability region which can be achieved by the FCSMA algorithm is a subject for future investigation.

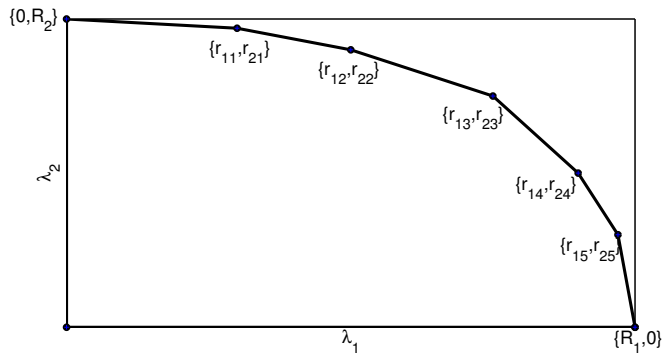

Figure 5.6: Stability region for a 2 user setup with multiple rate points.

5.8 Appendices

5.8.1 Proof of Proposition 6

Proof. The proof proceeds by considering a quadratic Lyapunov function of the form

$$V(\mathbf{Q}[t]) = \frac{1}{2}\sum_{i=1}^{2} Q_i^2[t]$$

and examining the value of λ for which the Lyapunov drift is negative outside a bounded set. Here, we only provide the essential technical arguments of the proof analyze the term

$$\dot{V}(\mathbf{Q}[t]) = \sum_{i=1}^{2} Q_i \dot{Q}_i[t], \qquad \dot{Q}_i[t] = \lambda - \mathbb{E}\left[S_i[t]\middle|\mathbf{Q}[t] = \mathbf{Q}\right]$$

which loosely represents the Lyapnov drift in continuous time. The exact analysis of the Lyapunov drift in discrete time is deferred till Section 5.8.2. We examine the range of λ for the which this quantity is negative.

$$\begin{aligned}\dot{V}(\mathbf{Q}[t]) &= \sum_{i=1}^{2} Q_i \dot{Q}_i[t] \\ &= \sum_{i=1}^{2} Q_i \left(\lambda - \frac{Q_i + \alpha^2 \sum_{k=1}^{2} Q_k}{\sum_{i=k}^{2} Q_k (1+\alpha)}\right)\end{aligned}$$

$$= \frac{\lambda(1+\alpha)\left(\sum_{i=1}^{2} Q_i\right)^2 - \left(\sum_{i=1}^{2} Q_i^2 + \alpha^2\left(\sum_{k=1}^{2} Q_k\right)^2\right)}{\sum_{i=k}^{2} Q_k(1+\alpha)}$$

$$= \frac{\lambda((1+\alpha)-\alpha^2)\left(\sum_{i=1}^{2} Q_i\right)^2 - \left(\sum_{i=1}^{2} Q_i^2\right)}{\sum_{i=k}^{2} Q_k(1+\alpha)} \tag{5.13}$$

$$\overset{(a)}{\leq} 0 \qquad \text{for} \qquad \frac{\alpha^2}{\alpha+1} \leq \lambda \leq \frac{\alpha^2}{\alpha+1} + \frac{1}{2(\alpha+1)} \tag{5.14}$$

where (a) follows from the following inequality. For $x, y, \beta_1, \beta_2 \geq 0$,

$$\begin{aligned}\beta_1(x+y)^2 - \beta_2(x^2+y^2) &= (\beta_1-\beta_2)x^2 + (\beta_1-\beta_2)y^2 + 2\beta_1 xy \\ &\leq (\beta_1-\beta_2)x^2 + (\beta_1-\beta_2)y^2 + \beta_1(x^2+y^2) \\ &= (2\beta_1-\beta_2)(x^2+y^2). \end{aligned} \tag{5.15}$$

Let us denote the numerator term of (5.13) as

$$g(\mathbf{Q}) \triangleq \lambda((1+\alpha)-\alpha^2)\left(\sum_{i=1}^{2} Q_i\right)^2 - \left(\sum_{i=1}^{2} Q_i^2\right).$$

The condition $\beta_1 > 0$ implies that $\lambda \geq \frac{\alpha^2}{\alpha+1}$. Note that $Q_1 \geq 0$ and $Q_2 \geq 0$. Rearranging the term inside the brackets of (5.13), it can be verified that

$$g(\mathbf{Q}) \leq 0 \text{ for } \lambda \leq \frac{\alpha^2}{\alpha+1} + \frac{1}{2(\alpha+1)}. \tag{5.16}$$

Combining with the condition $\lambda \geq \frac{\alpha^2}{\alpha+1}$, we have

$$g(\mathbf{Q}) \leq 0 \text{ for } \frac{\alpha^2}{\alpha+1} \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{1}{2(\alpha+1)}. \tag{5.17}$$

Notice that the range of λ specified in (5.16) is just a sufficient condition $g(\mathbf{Q}) \leq 0$. We now claim that

$$g(\mathbf{Q}) \leq 0 \qquad \text{for} \qquad 0 \leq \lambda \leq \frac{\alpha^2}{\alpha+1} + \frac{1}{2(\alpha+1)}. \tag{5.18}$$

We justify our claim in the following way. Let us define the upper bound on λ in (5.18) as $\lambda_{\max}$. Notice that $g(\mathbf{Q})$ is an increasing function of λ for a fixed value of the queue-lengths Q_1, Q_2. Therefore, $g(\mathbf{Q})\big|_{\lambda} \leq g(\mathbf{Q})\big|_{\lambda=\lambda_{\max}} \leq 0$ for $\lambda \leq \lambda_{\max}$ and hence the claim of (5.18).

Notice that the bound of (5.43) was obtained considering a fixed value of Q_1 and Q_2. However, the argument is true for any positive value of Q_1 and Q_2. Hence repeating the arguments for any Q_1 and Q_2, we conclude that the Lyapunov drift is negative for all positive values of queue-lengths and $\lambda < \lambda_{\max}$. We hence claim that the algorithm can stabilize the traffic whose arrival rate is less than $\lambda_{\max}$. The exact arguments and the connection to the Foster Lyapunov theorem is deferred till the proof of Theorem 2. □

5.8.2 Proof of Theorem 9

Proof. Consider the Lyapunov function given by

$$V(\mathbf{Q}[t]) = \frac{1}{2}\sum_{i=1}^{2}\left(Q_{ii}^2[t] + \delta Q_{i\bar{i}}^2[t]\right) \tag{5.19}$$

where $\bar{i} = \mod(i,2)+1$. Our approach to finding the maximum supportable rate is to examine the drift of the Lyapunov function and determine the maximum value of the arrival rate λ for which the Lyapunov drift is negative outside a bounded region around the origin. In doing so, we bound the Lyapunov function by a series of upper bounds and take the most restrictive condition on the arrival rate λ.

The Lyapunov drift in discrete time is given by

$$\Delta V(\mathbf{Q}[t]) = \mathbb{E}\left[V(\mathbf{Q}[t+1]) - V(\mathbf{Q}[t])|\mathbf{Q}[t] = \mathbf{Q}\right]$$

where $\mathbf{Q} = [Q_{11}, Q_{12}, Q_{21}, Q_{22}]^T$. Applying mean value theorem, considering $R_{ij}[t]$ between $Q_{ij}[t]$ and $Q_{ij}[t+1]$,

$$\begin{aligned}\Delta V(\mathbf{Q}[t]) &= \sum_{i=1}^{2}\mathbb{E}\left[R_{ii}[t](Q_{ii}[t+1]-Q_{ii}[t]) + \delta R_{i\bar{i}}[t]\left(Q_{i\bar{i}}[t+1]\right) - Q_{i\bar{i}}[t])\,\middle|\mathbf{Q}(t)=\mathbf{Q}\right] \\ &= \sum_{i=1}^{2}\mathbb{E}\left[R_{ii}[t](\dot{Q}_{ii}[t]+U_{ii}[t]) + \delta R_{i\bar{i}}[t](\dot{Q}_{i\bar{i}}[t]+U_{i\bar{i}}[t])\middle|\mathbf{Q}(t)=\mathbf{Q}\right]\end{aligned}$$

$$= \underbrace{\sum_{i=1}^{2}\mathbb{E}\left[R_{ii}[t]U_{ii}[t] + \delta R_{i\bar{i}}[t]U_{i\bar{i}}[t]\middle|\mathbf{Q}(t)=\mathbf{Q}\right]}_{\Delta V_1(\mathbf{Q}[t])} \tag{5.20}$$

$$+ \underbrace{\sum_{i=1}^{2}\mathbb{E}\left[R_{ii}[t]\dot{Q}_{ii}[t] + \delta R_{i\bar{i}}[t]\dot{Q}_{i\bar{i}}[t]\middle|\mathbf{Q}(t)=\mathbf{Q}\right]}_{\Delta V_2(\mathbf{Q}[t])} \tag{5.21}$$

Let us denote the term in equation (5.20) as $\Delta V_1(\mathbf{Q}[t])$ and (5.21) as $\Delta V_2(\mathbf{Q}[t])$. Consider

$$\Delta V_1(\mathbf{Q}[t]) = \sum_{i=1}^{2}\mathbb{E}\left[R_{ii}[t]U_{ii}[t] + \delta R_{i\bar{i}}[t]U_{i\bar{i}}[t]\middle|\mathbf{Q}(t)=\mathbf{Q}\right]. \tag{5.22}$$

We would like to bound the terms of $\Delta V_1(\mathbf{Q}[t])$. First note that if $Q_{ij}[t] = Q_{ij} > 1$ then $U_{ij}[t] = 0$. Else if $Q_{ij}[t] = Q_{ij} < 1$ and is selected for service then $0 < U_{ij} \leq 1$. In this case $Q_{ij}[t+1] < K+1$ (because $A_{ij}[t+1] < K$)and hence

$$\Delta V_1(\mathbf{Q}[t]) \leq b_1 K \tag{5.23}$$

where b_1 is a bounded positive constant. Now consider the terms of $\Delta V_2(\mathbf{Q}[t])$. Rewriting, we have,

$$\Delta V_2(\mathbf{Q}[t]) = \underbrace{\sum_{i=1}^{2} \mathbb{E}\left[R_{ii}[t]\dot{Q}_{ii}[t] + \delta R_{i\bar{i}}[t]\dot{Q}_{i\bar{i}}[t] \middle| \mathbf{Q}(t) = \mathbf{Q}\right] \mathbf{1}_{\mathbf{Q}\leq\mathbf{M}}}_{\Delta V_3(\mathbf{Q}[t])}$$
$$+ \underbrace{\sum_{i=1}^{2} \mathbb{E}\left[R_{ii}[t]\dot{Q}_{ii}[t] + \delta R_{i\bar{i}}[t]\dot{Q}_{i\bar{i}}[t] \middle| \mathbf{Q}(t) = \mathbf{Q}\right] \mathbf{1}_{\mathbf{Q}>\mathbf{M}}}_{\Delta V_4(\mathbf{Q}[t])}$$

where $\mathbf{1}_{(.)}$ is the indicator function. Note that since $A_{ij}[t] \leq K$, we can also conclude that $|A_{ij}[t] - S_{ij}[t]| \leq K$, therefore,

$$\Delta V_3(\mathbf{Q}[t]) \leq b_4(K+M)K. \tag{5.24}$$

In order to bound the terms of $\Delta V_4(\mathbf{Q}[t])$, first note that for a sufficiently large value of $Q_{ij}[t] = Q_{ij} > M$, we have $\left|\frac{R_{ij}[t]}{Q_{ij}} - 1\right| < \epsilon$ and therefore,

$$(1-\epsilon)Q_{ij} \leq R_{ij}[t] \leq (1+\epsilon)Q_{ij}.$$

Thus, we have

$$\begin{aligned}
&R_{ij}[t]\dot{Q}_{ij}[t] = R_{ij}[t]\left(A_{ij}[t] - S_{ij}[t]\right) \\
&= R_{ij}[t]\left((A_{ij}[t] - S_{ij}[t])_+ - (A_{ij}[t] - S_{ij}[t])_-\right) \\
&< (1+\epsilon)Q_{ij}(A_{ij}[t] - S_{ij}[t])_+ - (1-\epsilon)Q_{ij}(A_{ij}[t] - S_{ij}[t])_- \\
&= Q_{ij}(A_{ij}[t] - S_{ij}[t]) + \epsilon Q_{ij}[t]\left|A_{ij}[t] - S_{ij}[t]\right| \\
&\leq Q_{ij}\dot{Q}_{ii}[t] + \epsilon K Q_{ij}
\end{aligned} \tag{5.25}$$

where $(x)_+ = \max\{x, 0\}$, $(x)_- = -\min\{x, 0\}$ and $\left|A_{ij}[t] - S_{ij}[t]\right| \leq A_{ij}[t] \leq K$. Therefore, we have

$$\sum_{i=1}^{2} R_{ii}[t]\dot{Q}_{ii}[t] + \delta R_{i\bar{i}}[t]\dot{Q}_{i\bar{i}}[t] \leq \sum_{i=1}^{2} Q_{ii}\dot{Q}_{ii}[t] + \delta Q_{i\bar{i}}\dot{Q}_{i\bar{i}}[t] + K\epsilon\sum_{i=1}^{2}(Q_{ii} + \delta Q_{i\bar{i}}). \tag{5.26}$$

We focus on the first term on the right hand side of equation (5.26). Let us denote

$$\Delta V_5[t] \triangleq \sum_{i=1}^{2} \mathbb{E}\left[Q_{ii}\dot{Q}_{ii}[t] + \delta Q_{i\bar{i}}\dot{Q}_{i\bar{i}}[t] \middle| \mathbf{Q}(t) = \mathbf{Q}\right] \tag{5.27}$$

where

$$\Delta V_5[t] = Q_{11}\left(\mathbb{E}[A_{11}(t)] - \frac{Q_{11}}{B(\mathbf{Q})}\right) + \delta Q_{12}\left(\mathbb{E}[A_{12}(t)] - \frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) + Q_{22}\left(\mathbb{E}[A_{22}(t)] - \frac{Q_{22}}{B(\mathbf{Q})}\right) + \delta Q_{21}\left(\mathbb{E}[A_{21}(t)] - \frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right), \tag{5.28}$$

$B(\mathbf{Q}) = Q_{11} + \alpha(Q_{12}+Q_{21}) + Q_{22}$. Depending on the relationship between the queue-lengths, we need to consider the four cases for the Lyapunov function (see equation (5.8)). These four cases give rise to a series of bounds on λ under which the right hand side of equation (5.28) is negative. We take the most restrictive condition of all the bounds as the upper bound on the maximum supportable rate.

Case1: $Q_{11} \leq \delta Q_{12}; Q_{22} \leq \delta Q_{21}$.

In this case,

$$\begin{aligned}(5.28) =& Q_{11}\left(\lambda - \frac{Q_{11}}{B(\mathbf{Q})}\right) + \delta Q_{12}\left(\frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ &+ Q_{22}\left(\lambda - \frac{Q_{22}}{B(\mathbf{Q})}\right) + \delta Q_{21}\left(\frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ =& \frac{f_1(\mathbf{Q})}{B(\mathbf{Q})}\end{aligned}$$

where

$$f_1(\mathbf{Q}) = \lambda\left(Q_{11}+Q_{22}\right)\left(Q_{11}+\alpha(Q_{12}+Q_{21})+Q_{22}\right) - \left(Q_{11}^2 + \delta\alpha^2(Q_{12}+Q_{21})^2 + Q_{22}^2\right)$$

Let us examine the behavior of the function $f_1(\mathbf{Q})$ with respect to the variables Q_{12} and Q_{21} for a fixed value of Q_{11} and Q_{22}. Writing the gradients of the function $f_1(\mathbf{Q})$ with respect to the variables Q_{12} and Q_{21},

$$\begin{aligned}\frac{\partial f_1(\mathbf{Q})}{\partial Q_{12}} &= \alpha\lambda\left(Q_{11}+Q_{22}\right) - 2\alpha^2\delta(Q_{12}+Q_{21}) \\ &\leq \alpha\lambda\delta\left(Q_{12}+Q_{21}\right) - 2\alpha^2\delta(Q_{12}+Q_{21}) \\ &= Q_{12}\left(\alpha\lambda\delta - 2\alpha^2\delta\right) + Q_{21}\left(\alpha\lambda\delta - 2\alpha^2\delta\right) \\ &\leq 0 \qquad \text{for} \qquad \lambda \leq 2\alpha. \end{aligned} \tag{5.29}$$

Similarly taking the gradients with respect to Q_{21}, we have,

$$\frac{\partial f_1(\mathbf{Q})}{\partial Q_{21}} \leq 0 \qquad \text{for} \qquad \lambda \leq 2\alpha. \tag{5.30}$$

Therefore, $f_1(\mathbf{Q})$ is a decreasing function of both Q_{12} and Q_{21}. For a fixed value of Q_{11} and Q_{22}, the function $f_1(\mathbf{Q})$ is maximized when $\delta Q_{12} = Q_{11}$ and

$\delta Q_{21} = Q_{22}$ (hitting the boundary conditions of case 1). Therefore,

$$\begin{aligned} f_1(\mathbf{Q}) &\leq f_1(\mathbf{Q})\Big|_{\delta Q_{12}=Q_{11},\delta Q_{21}=Q_{22}} \qquad \text{for} \qquad \lambda \leq 2\alpha \\ &= \lambda(Q_{11}+Q_{22})\left(\left(1+\frac{\alpha}{\delta}\right)(Q_{11}+Q_{22})\right) - \left(Q_{11}^2+Q_{22}^2+\alpha^2\delta\left(\frac{(Q_{11}+Q_{22})^2}{\delta^2}\right)\right) \\ &= (Q_{11}+Q_{22})^2\left(\lambda\left(1+\frac{\alpha}{\delta}\right)-\frac{\alpha^2}{\delta}\right) - (Q_{11}^2+Q_{22}^2) \\ &\overset{(a)}{\leq} \left(2\left(\lambda\left(1+\frac{\alpha}{\delta}\right)-\frac{\alpha^2}{\delta}\right)-1\right)(Q_{11}^2+Q_{22}^2) \end{aligned} \tag{5.31}$$

where (a) follows from the inequality (5.15). The condition $\beta_1 \geq 0$ implies that $\lambda \geq \frac{\alpha^2}{\alpha+\delta}$. Note that $Q_{11} \geq 0$ and $Q_{22} \geq 0$. Rearranging the term inside the brackets of (5.31), it can be verified that

$$f_1(\mathbf{Q}) \leq 0 \text{ for } \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.32}$$

Combining with the condition $\lambda \geq \frac{\alpha^2}{\alpha+\delta}$, we have

$$f_1(\mathbf{Q}) \leq 0 \text{ for } \frac{\alpha^2}{\alpha+\delta} \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.33}$$

Notice that the range of λ specified in (5.32) is just a sufficient condition $f_1(\mathbf{Q}) \leq 0$. We now claim that

$$f_1(\mathbf{Q}) \leq 0 \qquad \text{for} \qquad 0 \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.34}$$

We justify our claim in the following way. Let us define the upper bound on λ in (5.34) as $\lambda_{\max}$. Notice the expression on the right hand side of equation (5.31) is an increasing function of λ for a fixed value of the queue-lengths Q_{11}, Q_{22}. Therefore, $f_1(\mathbf{Q})\big|_\lambda \leq f_1(\mathbf{Q})\big|_{\lambda=\lambda_{\max}} \leq 0$ for $\lambda \leq \lambda_{\max}$ and hence the claim of (5.34).

Also, it can be verified that

$$\frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)} \leq 2\alpha,$$

(the bound of $\lambda \leq 2\alpha$ is obtained from (5.30)). Hence, in this case, we have that for a given value of Q_{11} and Q_{22},

$$\Delta V_5[t] \leq 0 \qquad 0 \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.35}$$

Notice that the analysis of the result (5.35) was made assuming a fixed value of Q_{11} and Q_{22}. But note that the argument holds for any fixed values of Q_{11} and Q_{11}. Hence, repeating the argument for any Q_{11} and Q_{22}, we conclude that

(5.35) is negative for all values of positive value of queue-length and the range of λ specified.

Case2: $Q_{11} \geq \delta Q_{12}; Q_{22} \geq \delta Q_{21}$. In this case,

$$\begin{aligned}(5.28) =& Q_{11}\left(\frac{Q_{11}}{B(\mathbf{Q})}\right) + \delta Q_{12}\left(\lambda - \frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ &+ Q_{22}\left(\frac{Q_{22}}{B(\mathbf{Q})}\right) + \delta Q_{21}\left(\lambda - \frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ =& \frac{f_2(\mathbf{Q})}{B(\mathbf{Q})}\end{aligned}$$

where

$$f_2(\mathbf{Q}) = \lambda\delta\left(Q_{12}+Q_{21}\right)\left(Q_{11}+\alpha(Q_{12}+Q_{21})+Q_{22}\right) - \left(Q_{11}^2 + \delta\alpha^2(Q_{12}+Q_{21})^2 + Q_{22}^2\right).$$

Once again, we would like to examine the behavior of the function $f_2(\mathbf{Q})$ for fixed values of Q_{11} and Q_{22}. Writing the gradients of the function $f_2(\mathbf{Q})$ with respect to the variables Q_{12} and Q_{21},

$$\begin{aligned}\frac{\partial f_2(\mathbf{Q})}{\partial Q_{12}} &= \lambda\delta\left(Q_{11}+Q_{22}\right) + (2\alpha\lambda\delta - 2\alpha^2\delta)(Q_{12}+Q_{21}) \\ &\geq \lambda\delta^2\left(Q_{12}+Q_{21}\right) + (2\alpha\lambda\delta - 2\alpha^2\delta)(Q_{12}+Q_{21}) \\ &\geq 0 \qquad \text{for} \qquad \lambda \geq \frac{2\alpha^2}{2\alpha+\delta}. \end{aligned} \tag{5.36}$$

The function $f_2(\mathbf{Q})$ is an increasing function of Q_{12} for $\lambda \geq \frac{2\alpha^2}{2\alpha+\delta}$. Similarly, it can be shown that

$$\frac{\partial f_2(\mathbf{Q})}{\partial Q_{21}} \geq 0 \qquad \text{for} \qquad \lambda \geq \frac{2\alpha^2}{2\alpha+\delta}. \tag{5.37}$$

$f_2(\mathbf{Q})$ is an increasing function of both Q_{12} and Q_{21} for the range of λ specified. Therefore for a given value of Q_{11} and Q_{22}, the function $f_2(\mathbf{Q})$ is maximized when $\delta Q_{12} = Q_{11}$ and $\delta Q_{21} = Q_{22}$. Once again, repeating the arguments like that of case 1 (equation (5.31)), we have

$$f_2(\mathbf{Q}) \leq 0 \qquad \text{for} \qquad \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.38}$$

From the analysis of Case 2, there are two bounds on λ(from equations (5.36) and (5.38)). It can be verified that for $\delta \geq 0$,

$$\frac{2\alpha^2}{2\alpha+\delta} \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}$$

and hence,

$$\Delta V_5[t] \leq 0 \qquad \text{for} \qquad \frac{2\alpha^2}{2\alpha+\delta} \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.39}$$

Once again, we can make arguments similar to that of case 1 and extend the inequality in (5.39) for all positive values of the queue-lengths.

Case3: $Q_{11} \geq \delta Q_{12}; Q_{22} \leq \delta Q_{21}$.

In this case,

$$\begin{aligned}(5.28) =& Q_{11}\left(\lambda - \frac{Q_{11}}{B(\mathbf{Q})}\right) + \delta Q_{12}\left(\frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ &+ Q_{22}\left(\frac{Q_{22}}{B(\mathbf{Q})}\right) + \delta Q_{21}\left(\lambda - \frac{\alpha^2(Q_{12}+Q_{21})}{B(\mathbf{Q})}\right) \\ =& \frac{f_3(\mathbf{Q})}{B(\mathbf{Q})}\end{aligned}$$

where

$$f_3(\mathbf{Q}) = \lambda\left(Q_{11}+\delta Q_{21}\right)\left(Q_{11}+\alpha(Q_{12}+Q_{21})+Q_{22}\right) - \left(Q_{11}^2 + \delta\alpha^2(Q_{12}+Q_{21})^2 + Q_{22}^2\right).$$

Writing the gradients of the function $f_3(\mathbf{Q})$ with respect to the variables Q_{12} and Q_{21},

$$\begin{aligned}\frac{\partial f_3(\mathbf{Q})}{\partial Q_{12}} &= \lambda\delta Q_{11} + \lambda(\delta+\alpha)Q_{22} + (2\alpha\lambda\delta - 2\alpha^2\delta)Q_{12} + (\alpha\lambda\delta - 2\alpha^2\delta)Q_{21} \\ &\geq \lambda\delta^2 Q_{21} + \lambda(\delta+\alpha)\delta Q_{21} + (2\alpha\lambda\delta - 2\alpha^2\delta)Q_{12} + (\alpha\lambda\delta - 2\alpha^2\delta)Q_{21} \\ &\geq 0 \qquad \text{for} \qquad \lambda \geq \frac{2\alpha^2}{2\alpha+\delta}.\end{aligned}$$

Therefore, $f_3(\mathbf{Q})$ is an increasing function of Q_{12} for the range of λ specified and is maximized when $\delta Q_{12} = Q_{22}$. Examining the behavior of $f_3(\mathbf{Q})$ with respect to Q_{21},

$$\begin{aligned}\frac{\partial f_3(\mathbf{Q})}{\partial Q_{21}} &= (\alpha\lambda\delta - 2\alpha^2\delta)Q_{12} - 2\alpha^2\delta Q_{21} + \lambda\alpha Q_{22} \\ &\leq (\alpha\lambda\delta - 2\alpha^2\delta)Q_{12} + (\alpha\lambda\delta - 2\alpha^2\delta)Q_{21} \\ &\leq 0 \qquad \text{for} \qquad \lambda \leq 2\alpha.\end{aligned}$$

$$\frac{\partial f_3(\mathbf{Q})}{\partial Q_{21}} \leq 0 \qquad \text{for} \qquad \lambda \leq 2\alpha.$$

$f_3(\mathbf{Q})$ is a decreasing function of Q_{21} and hence is maximized when $\delta Q_{21} = Q_{22}$. Therfore for a given value of Q_{11} and Q_{22}, the function $f_3(\mathbf{Q})$ is maximized when $\delta Q_{12} = Q_{11}$ and $\delta Q_{21} = Q_{22}$.

$$f_3(\mathbf{Q}) \leq 0 \qquad \text{for} \qquad 0 \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.40}$$

Once again, we have that

$$\Delta V_5[t] \leq 0 \qquad \text{for} \qquad \frac{2\alpha^2}{2\alpha+\delta} \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.41}$$

Case4: $Q_{11} \geq \delta Q_{12}; Q_{22} < \delta Q_{21}$.

Case4 can be analyzed exactly similar to Case3. Once again it can be shown that the function is maximized when $\delta Q_{12} = Q_{11}$ and $\delta Q_{21} = Q_{22}$.

Summary: From the analysis of cases 1- 4, we have shown that

$$\Delta V_5[t] \leq 0 \qquad \text{for} \qquad \frac{2\alpha^2}{2\alpha+\delta} \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)} \tag{5.42}$$

for any positive values of the queue-lengths. Notice that the range of λ specified in (5.42) is just a sufficient condition for the Lyapunov drift to be negative. We now claim that

$$\Delta V_5[t] \leq 0 \qquad \text{for} \qquad 0 \leq \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.43}$$

We justify our claim in the following way. Notice that

$$\Delta V_5[t] = 0 \qquad \text{for} \qquad \lambda = \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}.$$

Also notice from equation (5.28) that $\Delta V_5[t]$ is an increasing function of λ for a fixed value of the queue-lengths $(Q_{11}, Q_{12}, Q_{21}, Q_{22})$. Hence

$$\Delta V_5[t]\Big|_{\lambda} \leq \Delta V_5[t]\Big|_{\lambda=\frac{\alpha^2}{\alpha+\delta}+\frac{\delta}{2(\alpha+\delta)}} \qquad \text{for} \qquad \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}. \tag{5.44}$$

This is true for a given value of queue-lengths. But recall that this argument can be extended for any value of the value of the queue-lengths. Therefore, (5.44) is true of all values of the queue-lengths and hence the claim of (5.43). Also note that

$$\Delta V_5[t] < 0 \qquad \text{for} \qquad 0 \leq \lambda < \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}.$$

In other words, for $\mathbf{Q} \in \mathbb{R}^+$ and $0 \leq \lambda < \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}$,

$$\lambda \left(Q_{11}\mathbf{1}_{Q_{11}\leq\delta Q_{12}} + \delta Q_{12}\mathbf{1}_{Q_{11}\geq\delta Q_{12}} + \delta Q_{21}\mathbf{1}_{Q_{22}\leq\delta Q_{21}} + Q_{22}\mathbf{1}_{Q_{22}\geq\delta Q_{12}}\right)$$
$$(Q_{11} + \alpha(Q_{12}+Q_{21}) + Q_{22}) - \left(Q_{11}^2 + \delta\alpha^2(Q_{12}+Q_{21})^2 + Q_{22}^2\right) < 0.$$

Denoting

$$g(\mathbf{Q}) \triangleq \left(Q_{11}\mathbf{1}_{Q_{11}\leq\delta Q_{12}} + \delta Q_{12}\mathbf{1}_{Q_{11}\geq\delta Q_{12}} + \delta Q_{21}\mathbf{1}_{Q_{22}\leq\delta Q_{21}} + Q_{22}\mathbf{1}_{Q_{22}\geq\delta Q_{12}}\right). \tag{5.45}$$

Therefore, there exists a $\eta > 0$ such that,

$$\left(Q_{11}^2 + \delta\alpha^2(Q_{12}+Q_{21})^2 + Q_{22}^2\right) \geq \lambda(1+\eta)g(\mathbf{Q})\left(Q_{11} + \alpha(Q_{12}+Q_{21}) + Q_{22}\right). \tag{5.46}$$

From (5.26) and (5.46), we have

$$\Delta V_4(\mathbf{Q}[t]) \leq (-\eta + K\epsilon)g(\mathbf{Q}). \tag{5.47}$$

By choosing ϵ sufficiently small, we can have $\gamma = \eta - K\epsilon > 0$, such that

$$\Delta V_4(\mathbf{Q}[t]) \leq -\gamma g(\mathbf{Q}). \tag{5.48}$$

Combining the result of (5.23),(5.24) and (5.48), we have

$$\Delta V(\mathbf{Q}[t]) = -\gamma g(\mathbf{Q})\mathbf{1}_{\mathbf{Q}>\mathbf{M}} + C \qquad \text{for} \qquad \lambda \leq \frac{\alpha^2}{\alpha+\delta} + \frac{\delta}{2(\alpha+\delta)}$$

where $C < \infty$ is a bounded positive constant. By Foster-Lyapunov theorem [103], implies that the queue-lengths are bounded for the FCSMA algorithm with $\lambda \leq \lambda_{\max}$. □

5.8.3 Proof of the Fading Channel Scenario

Proof. Consider the quadratic Lyapunov function given by

$$V(\mathbf{Q}[t]) = \sum_{i=1}^{2} \frac{1}{2}\left(Q_{ii}^2 + \delta Q_{i\bar{i}}^2[t]\right). \tag{5.49}$$

We once again analyze only the following expression of the Lyapunov drift given by

$$\begin{aligned}\dot{V}(\mathbf{Q}) = &\sum_{i=1}^{2} Q_{ii}\left(\mathbb{E}[A_{ii}(t)] - \sum_{s=1}^{|S|} p_s\left(\frac{R_s^2 Q_{ii}}{B_s(\mathbf{Q})}\right)\right)\\ &+ \delta Q_{i\bar{i}}\left(\mathbb{E}[A_{i\bar{i}}(t)] - \sum_{s=1}^{|S|} p_s\left(\frac{r_s^2\left(\sum_{k=1}^{2} Q_{k\bar{k}} + Q_{k\bar{k}}\right)}{B_s(\mathbf{Q})}\right)\right)\end{aligned} \tag{5.50}$$

where $B_s(\mathbf{Q}) = R_s Q_{11} + r_s(Q_{12} + Q_{21}) + R_s Q_{22}$. Let us denote the maximum supportable arrival rate in the fading case by the notation $\lambda_{\max}$. We will first analyze the Laypunov drift term at $\lambda = \lambda_{\max}$. Notice that $\lambda_{\max}$ can be written as a convex combination of λ^s (where λ^s are some rate points inside the stability region on the line $\lambda_1 = \lambda_2$) and hence $\lambda_{\max} = \sum_{s=1}^{|S|} p_s \lambda^s$. Therefore, we can rewrite the Lyapunov drift as

$$\begin{aligned}\dot{V}(\mathbf{Q}) = &\sum_{s=1}^{|\mathcal{S}|} p_s \sum_{i=1}^{2} \lambda^s (Q_{ii}\mathbf{1}_{Q_{ii}\leq\delta Q_{i\bar{i}}} + \delta Q_{i\bar{i}}\mathbf{1}_{Q_{ii}\geq\delta Q_{i\bar{i}}})\\ &- \left(\frac{R_s^2 Q_{ii} + r_s^2\left(\sum_{k=1}^{2} Q_{k\bar{k}} + Q_{k\bar{k}}\right)}{B_s(\mathbf{Q})}\right).\end{aligned} \tag{5.51}$$

From the proof of Theorem 2 , we have proved that each of the terms inside the summation for is negative (every channel state) as long as

$$\lambda^s < \frac{r_s^2}{r_s + \delta R_s} + \frac{\delta R_s^2}{2(r_s + \delta R_s)}$$

and hence from the above observation and $\lambda_{\max} = \sum_{s=1}^{|S|} p_s \lambda^s$, we have the result of (5.12). Also note that (5.50) is an increasing function of λ for a given value of queue-lengths. Hence, $\dot{V}(\mathbf{Q})\big|_{\lambda \leq \lambda_{\max}} < \dot{V}(\mathbf{Q})\big|_{\lambda = \lambda_{\max}}$ for $\lambda < \lambda_{\max}$. Also, this argument holds for any value of the queue-lengths. Therefore, $\dot{V}(\mathbf{Q}) \leq 0$ for $\lambda < \lambda_{\max}$ and for all values of queue-length and hence any rate $\lambda < \lambda_{\max}$ is stabilizable. □

Chapter 6

Conclusions & Outlook

Future cellular networks are forecasted to see an explosion in the wireless data traffic which will significantly stress the capacity of existing wireless networks. This has attracted people from the industry and academia alike to devise techniques which will enable wireless networks to sustain the growing data rate demand of the future.

The focus of this thesis is mainly the MIMO multi-cell setup. Most of the existing algorithms in MIMO multi-cell systems optimize the system parameters based on physical layer considerations. They perform resource allocation in a static manner and do not consider the stochastic nature of the traffic patterns. At the same time, there exist number of works in networking literature which focus on issues related to queuing stability and related concepts. However, most of these works simplify the modeling at the physical layer making it unrealistic for practical use. Therefore, new algorithms are needed which are able to take into account the randomness in the traffic and hence the queuing at the transmitter and at the same time account for physical layer parameters such as fading, pathloss, interference, imperfect CSI, power control, multiple antenna technologies etc. This thesis makes an attempt to incorporate higher layer processes such as traffic modeling, flow control in a MIMO multi-cell system. The two important considerations while developing these cross-layer design algorithms throughout this thesis have been the issue of decentralized design and energy efficiency.

We now briefly summarize the main results of this thesis and point out the possible limitations/ extensions in our system models and algorithms.

In Chapter 3, we formulated a reduced overhead beamforming algorithm using tools from RMT. Using asymptotic analysis, we proved the optimality of the reduced overhead beamforming strategy. We also derived in closed form the set of feasible target SINRs in a MIMO multi-cell scenario using large system

analysis. Our results show that exchanging information at the time scale of channel statistics yields good performance for practical system dimensions. A scope for future direction of this work could be to characterize the fluctuations of the downlink SINR for finite system dimensions. In order to compensate for the fluctuations of the achieved SINR around the target value, one could solve the problem with ROBF algorithm by considering a higher value target SINR (than the actual desired one).

Further in this chapter, we developed a flow controller for MIMO multi-cell systems using the H^{∞} control which functions without the knowledge of CSI of the wireless links. We would like to point out that in our solution, we decoupled the problem of beamforming and flow control into two separate parts. However, a better design method would be to jointly optimize the flow-controller and beamforming design from a cross-layer perspective (e.g. [13] develops a joint flow-controller/scheduling algorithm for a network with SISO links and conflict graph based interference model). The decoupling in our work can be seen as a likely sacrifice of performance in the interest of seeking decentralized solutions.

In Chapter 4, we considered the problem of minimizing the energy expenditure subject to long term QoS constraints in MIMO multi-cell systems. We showed that time average QoS constraints lead to greater energy savings as compared to instantaneous QoS constraints due to the flexibility they provide in dynamically allocating resources over fading channel states. Further, the solution developed was also decentralized in nature. A possible future direction of this work could be to design optimal beamforming vectors to satisfy time average rate constraints (of the form of $\log(1 + \text{SINR})$) and relate the problem to the stability of the real queue (as opposed to the virtual queue). Applying Lyapunov techniques to this problem leads to the weighted sum rate maximization problem in a MIMO multi-cell scenario which is known to be a non-convex optimization problem. Also, a possible extension is to consider the joint optimization of CSI feedback (optimizing the channel training during every time slot or optimizing the number of feedback bits in the case of quantization based feedback) and beamforming design.

Finally in Chapter 5, we have developed a FCSMA based distributed scheduling algorithm in an interference channel scenario under SINR based interference model. For analytical tractability, we considered a symmetric system set-up. Some of the possible extensions of the FCSMA algorithm to a general network set-up have been mentioned at the end of Chapter 5.

We finally provide some general conclusions derived from this thesis. The thesis was an attempt to incorporate issues of such as traffic flow and queuing in

MIMO multi-cell networks. From a physical layer point of view, our work provides a basic insight into the handling of randomness in arrival of information bits and the issue of queuing stability in MIMO systems. We show that performance improvements can be achieved by optimizing the system from a cross-layer perspective as compared to traditional physical layer based algorithms (e.g. in terms of average energy consumption). Our work also contributes to the issue of decentralized algorithm design which can approach the performance limits of the centralized algorithms. Our work also opens up new questions on how the networking based concepts can be broadened to handle physical layer issues such as fading, path-loss, interference, imperfect CSI and multiple antenna technologies.

We conclude the thesis by giving an outline of some interesting topics for future work:

6.1 Future Directions

6.1.1 CSI feedback and Queuing Stability for MIMO Multi-cell Networks

One of the fundamental bottlenecks in the performance of wireless systems is the problem of CSI acquisition. In particular, CSI feedback schemes have been widely investigate for multi-user MIMO [18, 19] (both analog and digital feedback) with the objective of maximizing metrics such as the ergodic sum rate. In this thesis, we have tried to relate the notion of queuing stability with multiple antenna communication techniques. A natural extension of our work is to relate queuing to CSI feedback schemes. The following questions can be addressed. How does the queuing stability region relate to the mean CSI feedback rates? Given an average CSI feedback rate, when and how accurately should the UTs feedback their CSI to the base station (BS)?

Another interesting problem could be to device joint CSI feedback and transmission techniques which achieve good performance with reduced CSI feedback rates. To this end, an interesting direction could be to exploit the time correlation inherent in the stochastic process associated with the channel. Most of the existing works use an i.i.d. process to model the channel state across time. The i.i.d. model has traditionally been a popular choice because of its simplicity and associated ease of analysis. On the other hand, this model fails to capture the temporal correlation inherent in realistic wireless channels. The time correlatedness of channel process has been modeled using Markov channels

techniques [20]. The memory inherent in the channel can help to reduce the CSI feedback load as compared to the i.i.d. models. The problem of joint optimization of transmission techniques and CSI feedback schemes can be modeled and solved using Markov decision processes (MDPs) similar to the works [21, 22]. Another interesting scenario is to optimize CSI feedback considering a channel model which is time-correlated but non-stationary which can be addressed using techniques such as learning algorithms.

6.1.2 Delay Efficient Algorithms

Novel multimedia applications such as video streaming, conferencing and gaming require require good guarantees in terms of delay performance. Delay-efficient algorithm design for queuing networks has also received attention in the networking community [23, 24]. It is well known that using multiple antennas at the transmitter provide additional degrees of freedom. An interesting question is how to exploit these additional degrees of freedom to provide better delay performance. One can also address the issue of exploiting the multiple antennas for deadline constrained traffic (which is critical in multimedia applications).

6.1.3 Opportunistic Resource Allocation Strategies for Cognitive Heterogeneous Networks

The ideas developed in this thesis can be extended to the set up of cognitive heterogeneous networks [25, 26, 27]. Cognitive heterogeneous networks refers to a cellular network with an underlay of an additional tier (e.g. femtocells) which is able to monitor and sense the environment and intelligently and dynamically allocate resources. In such networks, it is envisioned that the cognitive tier will be able to coexist and share the resources with the macro-cell by means of cognition and dynamic adaptation. In such a scenario, one can address the following question. Given a quality of service (QoS) constraint for the macro-cellular UTs, what is the maximum utility level at which the cognitive transmitter-receivers can co-exist. Alternately, how many cognitive transmitter receiver pairs can be simultaneously active with the macro cellular UTs. These problems can be formulated as stochastic optimization problems and one can devise on line algorithms using the technique of Lyapunov optimization. This scenario is particularly interesting since the asymmetry of data rate demands of the macro-cellular UTs, the randomness in the traffic arrival and the channel conditions will naturally provide a statistical multiplexing opportunity to enable cognitive transmitter receiver pairs to coexist. The cognitive transmitter

receiver communicating pairs can then opportunistically utilize the resources hence providing good throughput for both the tiers.

6.1.4 Optimal Transmission Techniques with Energy Harvesting

Future cellular networks will operate with small devices which are battery powered devices with limited battery capacity (such as small cell base stations). Energy harvesting has been identified as a viable solution in order to enable such networks to be operational for extended periods of time in a self-powered fashion. In networks having energy harvesting capability, energy becomes available for device transmissions at random times and in random amounts. In order to achieve good performance with energy harvesting systems, it is important to efficiently utilize the currently harvested energy and also the energy which will be harvested in future. One of the interesting future directions to look at would be to develop optimal transmission policies for future cellular networks with energy harvesting capability with the objective of optimizing metrics such as throughput, delay and achieving QoS guarantees. To this end, we can use techniques such as dynamic programming and Lyapunov optimization based approach [104, 105, 106].

Bibliography

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015," *White Paper*, Feb. 2011. viii, 2

[2] W. Webb, *Wireless Communications: The Future*. New York: Wiley, 2007. viii, 2

[3] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37 –43, march 2011. viii, ix, 2, 3, 21

[4] H. Claussen, "The Future of Small Cell Networks," *IEEE Comm. Soc.MMTC*, vol. 5, no. 5, pp. 32–36, Sep. 2010. viii, ix, 2, 3

[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005. ix, 3

[6] S. Sesia, I.Toufil, and E. M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 2nd ed. ix, 3

[7] G. J. Foschini and M. J. Gans, "On limits of Wireless Communications in a Fading Environment When Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998. ix, 3

[8] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999. ix, 3

[9] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Prentice Hall, 2007. ix, 3

[10] B. Rengarajan and G. de Veciana, "Architecture and Abstractions for Environment and Traffic-Aware System-Level Coordination of Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 721–734, Jun. 2011. ix, 3

[11] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992. x, 4, 87, 90

[12] A. Eryilmaz, R. Srikant, and J. Perkins, "Stable Scheduling Policies for Fading Wireless Channels," *IEEE/ACM Trans. Netw.*, vol. 13, no. 2, pp. 411–424, Apr. 2005. x, 4, 87, 90

[13] M. J. N. L. Georgiadis and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*. Now Publishers, 2006. x, xv, xvii, 4, 19, 20, 58, 61, 63, 112

[14] X. Lin, N. Shroff, and R. Srikant, "A Tutorial on Cross-layer Optimization in Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, Aug. 2006. x, 4

[15] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007. x, 4

[16] GreenTouch Consortium, "2010-2011 Annual report," whitepaper. xi, 5, 57

[17] G. Fettweis and E. Zimmermann, "ICT Energy Consumption - Trends and Challenges," in *The 11th International Symposium on Wireless Personal Multimedia Communications (WPMC'08)*, 2008. xi, 5, 57

[18] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006. xix, 113

[19] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010. xix, 113

[20] Q. Zhang and S. Kassam, "Finite-State Markov Model for Rayleigh Fading Channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999. xix, 114

[21] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. Shroff, "Exploiting Channel Memory for Joint Estimation and Scheduling in Downlink Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'11)*, Apr. 2011, pp. 3056–3064. xx, 114

[22] W. Ouyang, A. Eryilmaz, and N. Shroff, "Asymptotically Optimal Downlink Scheduling over Markovian Fading Channels," in *IEEE International Conference on Computer Communications, (INFOCOM'12)*, Apr. 2012. xx, 114

[23] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, "Scheduling for Small Delay in Multi-rate Multi-channel Wireless Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'11)*, Apr. 2011, pp. 1251–1259. xx, 114

[24] L. Huang and M. Neely, "Delay Reduction via Lagrange Multipliers in Stochastic Network Optimization," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 842–857, Apr. 2011. xx, 114

[25] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femtocells: Past, Present, and Future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012. xx, 114

[26] A. Adhikary, V. Ntranos, and G. Caire, "Cognitive Femtocells: Breaking the Spatial Reuse Barrier of Cellular Systems," in *Information Theory and Applications Workshop (ITA'11)*, Feb. 2011, pp. 1–10. xx, 114

[27] A. Adhikary and G. Caire, "On the Coexistence of Macrocell Spatial Multiplexing and Cognitive Femtocells," in *1st International Workshop on Small Cell Wireless Networks (SmallNets), ICC'12*, Jun. 2012. xx, 114

[28] J. E. Rombaldi, *Analyse Matricielle*. EDP Sciences, 2000. 10

[29] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985. 10

[30] J. W. Silverstein and Z. Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995. 10, 12, 14

[31] P. Billingsley, *Probability and Measure*. John Wiley & Sons, Inc., Hoboken, NJ, 1995. 11, 49

[32] Z. Bai and J. W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 26, pp. 316–345, Jan. 1998. 11

[33] M. Peacock, I. Collings, and M. Honig, "Eigenvalue Distributions of Sums and Products of Large Random Matrices Via Incremental Matrix Expansions," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, May 2008. 12

[34] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice Hall, 2000. 15

[35] T. Basar and P. Bernhard, *H-∞ Optimal Control and Relaxed Minimax Design Problems: A Dynamic Game Approach*. Birkhauser, Boston, MA, 1991. 16, 41

[36] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO Paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007. 26

[37] M. Karakayali, G. Foschini, and R. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, Aug. 2006. 26

[38] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming Intercell Interference in Indoor Wireless Systems," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC'07)*, Nov. 2007, pp. 83–87. 26

[39] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010. 26

[40] S. Shamai and B. Zaidel, "Enhancing the Cellular Downlink Capacity via Co-Processing at the Transmitting End," in *IEEE Vehicular Technology Conference (VTC'01)*, vol. 3, 2001, pp. 1745–1749. 26

[41] H. Dahrouj and W. Yu, "Coordinated Beamforming for the Multi-Cell Multi-Antenna Wireless System," in *Conference on Information Sciences and Systems, (CISS'08), Princeton, NJ, USA*, Mar. 2008, pp. 429–434. 26

[42] ——, "Coordinated Beamforming for the Multicell Multi-antenna Wireless System," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010. 26, 27, 28, 29, 30, 47

[43] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Distributed Multicell and Multiantenna Precoding: Characterization and Performance Evaluation," in *Proc. Global Communications Conference (GLOBECOM'09)*, Dec. 2009, pp. 1–6. 26

[44] E. Larsson and E. Jorswieck, "Competition Versus Cooperation on the MISO Interference Channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008. 26

[45] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete Characterization of the Pareto Boundary for the MISO Interference Channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008. 26

[46] R. Zakhour and D. Gesbert, "Coordination on the MISO Interference Channel Using the Virtual SINR Framework," in *Proc. International ITG Workshop on Smart Antennas*, Feb. 2009. 26

[47] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality Properties, Distributed Strategies, and Measurement-Based Evaluation of Coordinated Multicell OFDMA Transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011. 26

[48] B. L. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed Downlink Beamforming With Cooperative Base Stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, Dec. 2008. 27

[49] R. Zhang and S. Cui, "Cooperative Interference Management With MISO Beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010. 27

[50] D. Tse and S. Verdu, "Optimum Asymptotic Multiuser Efficiency of Randomly Spread CDMA," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2718–2722, Nov. 2000. 27

[51] D. Tse and S. Hanly, "Linear Multiuser Receivers: Effective Interference, Effective Bandwidth and User Capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999. 27

[52] D. Tse and O. Zeitouni, "Linear Multiuser Receivers in Random Environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000. 27

[53] A. Tulino, A. Lozano, and S. Verdu, "Impact of Antenna Correlation on the Capacity of Multiantenna Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005. 27

[54] R. Couillet, M. Debbah, and J. Silverstein, "A Deterministic Equivalent for the Analysis of Correlated MIMO Multiple Access Channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011. 27

[55] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An

Asymptotic Approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1048–1069, Mar. 2010. 27

[56] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal Channel Training in Uplink Network MIMO Systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Jun. 2011. 27

[57] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012. 27

[58] R. Couillet and M. Debbah, *Random Matrix Theory Methods for Wireless Communications*. Cambridge University Press, 2011. 27, 47

[59] S. Lakshminaryana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," in *International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'10)*, Sep. 2010, pp. 2105–2110. 27

[60] R. Zakhour and S. Hanly, "Base Station Cooperation on the Downlink: Large System Analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012. 27, 35

[61] A. Wiesel, Y. Eldar, and S. Shamai, "Linear Precoding via Conic Optimization for Fixed MIMO Receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006. 28, 35

[62] R. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995. 43

[63] W. Hachem, P. Loubaton, and J. Najim, "Deterministic Equivalents for Certain Functionals of Large Random Matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007. 48

[64] J. Zander and M. Frodigh, "Comment on "Performance of optimum transmitter power control in cellular radio systems"," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, p. 636, Aug. 1994. 52

[65] T. C. Group and G. e Sustainability Initiative (GeSI), "SMART 2020: Enabling the Low Carbon Economy in the Information Age." [Online]. Available: http://www.smart2020.org/ 57

[66] E. Belmega and S. Lasaulce, "Energy-Efficient Precoding for Multiple-Antenna Terminals," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 329–340, Jan. 2011. 57

[67] S. Cui, A. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004. 57

[68] H. Kim, C.-B. Chae, G. de Veciana, and R. Heath, "A Cross-Layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009. 57

[69] H. S. Kim and B. Daneshrad, "Energy-Constrained Link Adaptation for MIMO OFDM Wireless Communication Systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2820–2832, Sep. 2010. 57

[70] Z. Chong and E. Jorswieck, "Energy Efficiency in Random Opportunistic Beamforming," in *Proc. of IEEE 73rd Vehicular Technology Conference, Budapest, Hungary (VTC Spring)*, 2011. 57

[71] ——, "Energy-Efficient Power Control for MIMO Time-Varying Channels," in *IEEE Online Conference on Green Communications (GreenCom)*, Sep. 2011, pp. 92–97. 57

[72] R. Zhang, "On Peak versus Average Interference Power Constraints for Protecting Primary Users in Cognitive Radio Networks," *IEEE Trans. Signal Process.*, vol. 8, no. 4, pp. 2112–2120, Apr. 2009. 57

[73] E. Nekouei, H. Inaltekin, and S. Dey, "Throughput Scaling in Cognitive Multiple-Access with Average Power and Interference Constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 927–946, Feb. 2012. 57

[74] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, and J. Ramos, "Resource Allocation for Interweave and Underlay CRs under Probability-of-Interference Constraints." [Online]. Available: http://www.dtc.umn.edu/s/resources/jsac2012antonio.pdf 57

[75] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energy-Efficient Packet Transmission over a Wireless Link," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 487–499, Aug. 2002. 57

[76] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002. 57

[77] M. Neely and R. Urgaonkar, "Opportunism, Backpressure, and Stochastic Optimization with the Wireless Broadcast Advantage," in *42nd Asilomar Conference on Signals, Systems and Computers (ASILOMAR'08)*, Oct. 2008, pp. 2152–2158. 58

[78] M. Neely, "Energy Optimal Control for Time-Varying Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006. 58, 81

[79] M. Gastpar, "On Capacity Under Receive and Spatial Spectrum-Sharing Constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 471–487, Feb. 2007. 60

[80] R. Zhang, Y. Liang, and S. Cui, "Dynamic Resource Allocation in Cognitive Radio Networks ," *IEEE Signal Process. Mag.*, vol. 5, no. 27, pp. 102–114, May 2010. 60

[81] A. El-Keyi and B. Champagne, "Cooperative MIMO-Beamforming for Multiuser Relay Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, 2008. 60

[82] M. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010. 62, 63

[83] J. F. Sturm, "Using SEDUMI 1.02, a Matlab Toolbox for Optimizations Over Symmetric Cones," *Optimization Meth. and Soft*, vol. 11-12, May 1999. 67

[84] G. Pataki, "On the Rank of Extreme Matrices in Semidefinite Programs and the Multiplicity of Optimal Eigenvalue," *Math. Oper. Res.*, vol. 23, no. 2, pp. 339–358, 1998. 67

[85] B. Hassibi and B. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003. 73

[86] B. Li and A. Eryilmaz, "A Fast-CSMA Algorithm for Deadline-Constrained Scheduling over Wireless Fading Channels," in *RAWNET/WNC3, 9th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, May 2011. 86, 95

[87] L. Tassiulas and A. Ephremides, "Dynamic Server Allocation to Parallel Queues with Randomly Varying Connectivity," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 466–478, Mar. 1993. 87

[88] S. Shakkottai and A. L. Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule," *American Mathematical Society Translations, Series*, vol. 2, p. 2002, 2000. 87

[89] M. Neely, E. Modiano, and C. Rohrs, "Dynamic Power Allocation and Routing for Time Varying Wireless Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'03)*, vol. 1, Apr. 2003, pp. 745–755. 87

[90] L. Tassiulas, "Linear Complexity Algorithms for Maximum Throughput in Radio Networks and Input Queued Switches," in *IEEE International Conference on Computer Communications, (INFOCOM'98)*, vol. 2, Mar. 1998, pp. 533–539. 87

[91] A. Eryilmaz, A. Ozdaglar, D. Shah, and E. Modiano, "Distributed Cross-Layer Algorithms for the Optimal Control of Multihop Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 63–651, Apr. 2010. 87

[92] L. Bui, S. Sanghavi, and R. Srikant, "Distributed Link Scheduling With Constant Overhead," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1467–1480, Oct. 2009. 87

[93] C. Joo, X. Lin, and N. Shroff, "Understanding the Capacity Region of the Greedy Maximal Scheduling Algorithm in Multi-Hop Wireless Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'2008)*, Apr. 2008, pp. 1103–1111. 87

[94] L. Bui, A. Eryilmaz, R. Srikant, and X. Wu, "Joint Asynchronous Congestion Control and Distributed Scheduling for Multi-Hop Wireless Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'06)*, Apr. 2006, pp. 1–12. 87

[95] X. Lin and N. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Wireless Networks," in *IEEE International Conference on Computer Communications, (INFOCOM'05)*, vol. 3, Mar. 2005, pp. 1804–1814. 87

[96] L. Jiang and J. Walrand, "A distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks," in *Annual Allerton Confer-*

ence on Communication, Control, and Computing, (Allerton'08), Sep. 2008, pp. 1511–1519. 87

[97] ——, "A Distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, Jun. 2010. 87, 93

[98] S. Rajagopalan, D. Shah, and J. Shin, "Network Adiabatic Theorem: an Efficient Randomized Protocol for Contention Resolution," in *Proc. ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'09)*, 2009, pp. 133–144. 87

[99] J. Ghaderi and R. Srikant, "On the Design of Efficient CSMA Algorithms for Wireless Networks," in *IEEE Conference on Decision and Control (CDC'10)*, Dec. 2010, pp. 954–959. 87

[100] B. Rong and A. Ephremides, "Joint MAC and Rate Control for Stability and Delay in Wireless Multi-Access Channels," *Perform. Eval.*, vol. 68, pp. 658–669, Aug. 2011. 88

[101] D. Qian, D. Zheng, J. Zhang, and N. Shroff, "CSMA-Based Distributed Scheduling in Multi-hop MIMO Networks under SINR Model," in *IEEE International Conference on Computer Communications, (INFOCOM'10)*, Mar. 2010, pp. 1–9. 90

[102] J. Jose and S. Vishwanath, "Distributed Rate Allocation for Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6539–6554, Oct. 2011. 90

[103] S. Asmussen, *Applied Probability and Queues*. New York, NY, USA: Springer-Verlag, 2003. 109

[104] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Control of Wireless Networks with Rechargeable Batteries," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 581–593, Feb. 2010. 115

[105] K. Tutuncuoglu and A. Yener, "Optimum Transmission Policies for Battery Limited Energy Harvesting Nodes." [Online]. Available: http://arxiv.org/abs/1010.6280 115

[106] L. Huang and M. J. Neely, "Utility Optimal Scheduling in Energy Harvesting Networks," in *Proc.of Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'11)*, May 2011. 115